STRUCTURAL ORIGINS OF PRESSURE EFFECTS IN PROTEINS

by
José Alfredo Caro

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of PhilosophyJohns Hopkins University

Baltimore, Maryland

February, 2015

Abstract

The molecular mechanisms of chemical and heat denaturation of proteins are relatively well established; those of pressure unfolding are not. Volume is the conjugate variable of pressure; it is the fundamental thermodynamic variable that governs the pressure sensitivity of proteins. Cavities that are present in the native state and absent in the unfolded state are thought to contribute significantly to the change in volume upon unfolding (ΔV). Staphylococcal nuclease was used to examine the role of cavities systematically. The wild-type protein has a small cavity in its hydrophobic core, comparable in volume to a water molecule. Artificial cavities were generated by substitution of internal hydrophobic residues to Ala. Substitutions of small residues with large ones were used to eliminate the natural cavity. Substitutions to polar residues were used to affect the hydration state of cavities. For 27 variants studied, (a) crystal structures, (b) thermodynamic stabilities using chemical denaturation, and (c) ΔV of unfolding measured by pressure denaturation monitored with Trp-fluorescence and NMR spectroscopy were obtained. In general, the cavities did not affect the structure. The cavities were large enough to hold several waters, but these were only detected in the cavities lined with polar groups. The measured ΔV of variants was always larger than for the wild-type. A near-linear correlation between the ΔV measured experimentally and the one calculated from structures illustrate the importance of cavities in pressure sensitivity. A correlation between measured ΔV and thermodynamic stability (ΔG°) suggests that 1 kcal/mol is lost per 11 mL/mol of increased void volume. This study demonstrates that cavities contribute significantly towards the pressure sensitivity of proteins and can modulate the hydration and structural fluctuations of proteins.

Thesis Committee

Advisor:	Dr. Bertrand García-Moreno E.
Reader:	Dr. Juliette Lecomte
Committee:	Dr. Vincent Hilser
	Dr. L. Mario Amzel
	Dr. Richard Cone

Acknowledgements

I am eternally grateful for the unwavering support from my advisor Dr. Bertrand García-Moreno E. and the boundless insight provided during the course of my research. Likewise, I am forever indebted to Dr. Catherine Ann Royer for her support in this venture. I am thankful to the members of the Jenkins and of the Program of Molecular Biophysics faculty and in particular to the members of my committee, Dr. Vincent Hilser, Dr. Richard Cone and Dr. L. Mario Amzel, for their devotion to science and education that has brought me here. In particular, I would like to thank Dr. Juliette Lecomte for being a role model to me and for her insightful comments and challenging questions, an intellectual space that will be difficult to fill. I am grateful to Dr. Jamie Schlessman for the patient training and support, to Dr. Angel E. García, Dr. Mariano Dellarole, Dr. Julien Roche, Dr. Jean-Baptiste Rouget and Martin Fossat, for the collaborations, and to Dr. Carolyn Fitch and the members of the García-Moreno lab for maintaining my own and the laboratory's health and bustling activity. I thank Dr. Robert Trachman, Dr. Juan Perilla, Dr. Marcin Balicki, Hesam Motlag, the finished class of 2008, Ranice Crosby, Jessica Appel, Alexias Ebert and the Jenkins administrative and technical staff. I leave blessed with their friendship. Most of all, I thank my parents for their love, my sister for the reference state, Bodie and Hagebutte for being fierce but fluffy, and the wonderful, ever-glowing Hannah Aliza Cohen.

Table of Contents

List of tables

List of figures

CHAPTER 1:

STRUCTURAL ORIGINS OF PRESSURE EFFECTS IN PROTEINS

1. Introduction

In 1914 Percy W. Bridgman used raw egg white to seal a pressure gauge and haphazardly discovered that albumen coagulates under pressure into a state resembling "a hard boiled egg (1)." He discovered that pressure can unfold proteins. Pressure and volume being conjugate thermodynamic variables, it follows from Le Châtelier's principle that Bridgman's discovery implied that the volume of the folded state must be greater than that of the denatured state. One hundred years later, the structural origin of this difference in volume (ΔV) remains poorly understood. That is the main topic of this dissertation.

In 1970, Brandts et al. explained that the general principles used to describe the effects of heat on protein stability fail to describe how pressure denatures proteins (2). The transfer of non-polar, model compound amino acids from a liquid hydrocarbon into water can be used to correlate the change in accessible surface area upon unfolding with the experimentally measured heat capacity (3). In contrast, the measured ΔV of proteins cannot be predicted using transfer volumes of non-polar compounds and does not correlate with the change in surface area upon unfolding (4–7).

Here I will review a number of possible factors that have been identified by previous research as being important determinants of the pressure sensitivity of proteins. Internal cavities present in the native state and absent in the unfolded state are thought to make a key contribution to ΔV and to the pressure sensitivity of proteins. That is the main idea examined by the experiments discussed in this dissertation.

1. Pressure unfolds proteins

Pressure and temperature are the two fundamental thermodynamic variables that govern the properties of biological macromolecules. Studies of pressure perturbation of proteins contribute a wealth of insight that cannot be accessed using other perturbants such as temperature, pH, or chemical denaturants or osmolytes. For example, pressure reveals volumetric properties of a system without altering the solution composition. Perturbation with pressure is usually fully reversible and it can even be used to reverse seemingly irreversible reactions (i.e., pressure can dissociate aggregates) (8, 9). The pressure unfolded state of proteins retains residual structure that is absent in a heat or chemically denatured state (10–12). Perhaps most importantly, unlike the response of proteins to temperature, which is global, the response of proteins to pressure is local and therefore it is heterogeneous (i.e., different regions of a protein are affected differently by pressure). This allows stabilization of low-lying excited states that are suppressed by stronger denaturants such as urea or heat, and therefore invisible (11–13). High pressures also enable studies of protein folding a low temperatures, providing further insight into the phase diagram of proteins (14). Lastly, the volumetric fluctuations induced by pressure hold information on the hydration state of the protein and can induce the hydration of seemingly hydrophobic environments such as the protein core (15, 16).

1.2 Thermodynamics of pressure unfolding of proteins

In 1971, Hawley implemented a thermodynamic theory that unified the behavior of proteins as a function of both pressure and temperature (17). According to his theory,

the free energy of folding, ΔG, of a protein at p_0 and T_0 in equilibrium between two states is described by

$$\Delta G(p,T) = \Delta G°(p_0,T_0) + \Delta V°(p-p_0) + \frac{1}{2}\Delta K°(p-p_0)^2 + \Delta E°(p-p_0)(T-T_0)$$

$$- \Delta C_p°[T(ln\frac{T}{T_0} - 1) + T_0] - \Delta S°(T-T_0) \qquad (1)$$

ΔG° describes the change in Gibbs energy of the system at p_0 and T_0 with respect to the standard state (1 bar). The change in heat capacity, $\Delta C_p°$ and the change in entropy, ΔS°, account for the temperature dependence. The pressure dependence involves the change in volume, ΔV°, and the change in compressibility, ΔK°. The change in expansivity, ΔE°, depends on both the temperature and the pressure. The range of pressure and temperature typically accessible experimentally precludes any accurate measurement of terms of order higher than quadratic and any mechanistic understanding of the response of proteins to pressure (2, 4, 18–20). Applying equation (1) to proteins assumes that the temperature- and pressure-induced unfolded states are equivalent. In other words, the free energy of unfolding by pressure is equal to that of the unfolding by temperature. However, seminal work by the Jonas laboratory showed that the heat denatured state is significantly less structured than the pressure unfolded state (11). In fact, cold-denatured ubiquitin at 2.5 kbar resembles the partially folded A-state of ubiquitin induced by alcohol (12). Nevertheless, equation (1) serves as a first approximation of the phase diagram of proteins.

1.3 Expansivity

In equation (1), the coupling between pressure and temperature is treated explicitly by the expansivity term, defined as

$$\frac{\partial^2 \Delta G(p,T)}{\partial p \partial T} = \left(\frac{\partial \Delta V}{\partial T}\right)_P = V \Delta \alpha^\circ = \Delta E^\circ \qquad (2)$$

Pressure perturbation calorimetry (PPC) offers an approach for direct measurement of the expansivity of proteins and of small organic molecules (20). The seminal work by the Brandts laboratory showed that the expansivity at room temperature of polar and non-polar side chains were of opposite sign. The polar side chains, which break the structure of bulk water, reported positive expansivities, while the opposite was true for the structure-making non-polar side chains. It is intuitive to think that a solvent that becomes structured in the presence of a non-polar side chain will reduce its molar volume as temperature is increased and kinetic energy overcomes the need to satisfy hydrogen bonds. Indeed, as the temperature was increased or when large amounts of Gdn_2SO_4 were present, the expansivity values for the side chain converged to the values of pure water. This indicates that, under such large perturbations, the different hydration properties of individual amino acids are lost. The Royer and Winter groups extended the work with pressure perturbation calorimetry by studying tripeptides and amino acids with blocked termini (21). They found that the structural context of the side chain can greatly influence its expansivity. A hydrophobic side chain in a tripeptide exhibits a less negative expansivity, i.e. a less hydrophobic character, than as a single amino acid, even after subtraction of the reference tripeptide (21, 22).

1.3.1 Change in expansivity upon protein unfolding

To evaluate the contribution from expansivity to protein stability it is necessary to measure the change in expansivity upon unfolding by measuring the expansivity of both the folded and unfolded states. The expansivity of folded proteins measured by pressure perturbation calorimetry near room temperature generally falls somewhere between the values measured for polar and non-polar amino acids, suggesting a balanced contribution from its side chains (20, 23). Conversely, the expansivity of a constitutively unfolded variant of SNase, T62P, was measured near room temperature and used as an approximation of the expansivity of the unfolded state of SNase. Interestingly, the T62P variant shows an expansivity comparable to that of the wild type protein (24). This suggests that the change in expansivity upon unfolding of proteins might be negligible.

At constant temperature, the expansivity term vanishes from equation (1). In pressure denaturing experiments of proteins, however, a contribution from expansivity can still exist. For example, if the pressure unfolding of a protein is being compared to that of its variant with a single point mutation, then the change in expansivity between the two proteins needs to be taken into account as a possible contribution to ΔV. In a study by Dellarole et al., when 11 surface residues in BPTI (5 ionizable, 3 polar, 1 hydrophobic and 2 glycine residues) were substituted simultaneously to Ala, no significant change in expansivity was measured. In contrast, variants of SNase with a single Ala substitution in the core doubled the expansivity of SNase (from α = 450 1/K to α = 1200 1/K for variant V66A, measured at 10°C in terms of the coefficient of thermal expansion $\alpha = E/V$, where V is the molar volume of the protein) (24). These variants also reported increases in ΔV between 12 mL/mol and 55 mL/mol relative to the ΔV of SNase (58 mL/mol). If one

assumes that the substitutions had no effect on the expansivity of the unfolded state, then the change in expansivity, ΔE, between SNase and its variants could give rise to the measured changes in ΔV. However, no correlation between the increases in expansivity and the increases in ΔV was observed, suggesting that expansivity cannot be the dominant contribution to ΔV (13, 24).

1.4 Compressibility

The pressure dependence of the free energy is given by

$$\left(\frac{\partial \Delta G(p)}{\partial p}\right)_T = \Delta V(p) \qquad (2)$$

A Taylor series expansion around pressure p_0 yields the explicit dependence of ΔV on pressure.

$$\Delta V(p) = \Delta V(p_0) + \left.\frac{d\Delta V(p)}{dp}\right|_{p_0} (p - p_0) + \cdots \qquad (3)$$

The derivative of ΔV with respect to pressure is the isothermal compressibility, ΔK. By choosing p_0 to be the standard state pressure of 1 bar, the integral of equation (2) yields

$$\Delta G(p) = \Delta G° + \Delta V°(p - p_0) + \frac{1}{2}\Delta K°(p - p_0)^2 + \cdots \qquad (4)$$

The experimental measurements of the magnitude of the change in isothermal compressibility upon protein unfolding report conflicting results (2, 4, 17, 25–30). A direct measurement of isothermal compressibility can be done by densitometry. The Royer laboratory made the first measurement of the change in isothermal compressibility of a protein, SNase, by measuring the change in volume as a function of pressure of

native SNase and its thermally unfolded form (25). They found that the measured compressibility for the folded state of SNase is indistinguishable from that of the unfolded state. If this is the case, equation (4) can be reduced to

$$\Delta G(p) = \Delta G° + \Delta V°(p - p_0) \qquad (5)$$

However, seminal work by the Jonas laboratory showed that the heat denatured state is significantly less structured than the pressure unfolded state, and might not be representative of the change in compressibility upon pressure unfolding (11). In addition, it is highly likely that the mechanisms of compression of the folded and unfolded states of proteins are different (31). While the homogeneous response of the hydration layer is expected to dominate the compressibilitiy of the unfolded state (32, 33), the folded state showed a location dependent response to pressure by a number of different techniques, reflecting perhaps the heterogeneous distribution of cavities or of the structural fluctuations that gives rise to cavities (16, 34–37). If this is indeed the case, different regions of the protein should experience different changes in compressibility upon unfolding, which in turn will represent a determining factor in the pressure stability of proteins.

1.4.1 Effects of pressure on the native state of proteins

High pressure crystallography enables the study of the behavior of proteins at increased pressures with atomic resolution. As early as 1987, the Richards laboratory (38) monitored the effect of pressure (1000 bar) on crystals of hen egg white lysozyme and already ascribed different compressibility values to the two domains of the protein. More recently, the Watanabe laboratory (35) identified the general reduction of volume of most

internal cavities of 3-isopropylmalate dehydrogenase, and a single cavity that retained its volume even at pressures as high as 6500 bar. This cavity also showed an increase in electron density, which the authors ascribed to new, buried water molecules.

It is noteworthy that, for a protein constrained in a crystal lattice, some cavities shrink as pressure increases while others retain the same volume observed at 1 bar. At very high pressures of 6 kbar, the cavities that retain their volume show increased evidence of internal waters, and that might be the reason that their volume is unchanged. Much remains to be understood about the interplay between the compression of a cavity and the preservation of its volume under pressure, either because of the rigidity imparted by other protein interactions or because of pressure-induced water penetration. High pressure NMR spectroscopy can provide insight into the dynamics that determine such volumetric fluctuations. Early work by the Jonas (39), the Wand (40) and the Akasaka (27) laboratories implemented different aspects of NMR spectroscopy to reveal a highly heterogeneous response of proteins to pressure.

The Jonas laboratory monitored proton resonances of select amino acids in hen egg white lysozyme and found that the measured intensities decreased heterogeneously. The Wand laboratory found that the protection factors of apocytochrome b_{562} obtained from hydrogen exchange experiments as a function of pressure were consistent with the cooperative subunits they identified using hydrogen exchange experiments as a function of GdnHCl (41). The Akasaka laboratory monitored the chemical shifts of amide resonances of bovine pancreatic trypsin inhibitor (BPTI) as a function of pressure and interpreted the heterogeneous shifts as a loss of structure localized to the most pressure sensitive regions of the protein. The non-linear changes in chemical shifts as a function of

pressure were interpreted as the population of low-lying excited states (28). The same laboratory would later publish the first high pressure (3000 bar) NMR structure of a protein, ubiquitin, showing structural deviations with respect to the NMR structure at 30 bar localized to the helical region of the protein (42).

Pressure induced changes in the native structure and shifts in population towards alternative conformations represent possible mechanisms of compression. The experiments listed above are some of the earliest work describing the local nature of the compressibility of native proteins. They suggest that the change in compressibility upon unfolding is region-specific, and represents a significant contribution to the pressure stability of different regions of the protein. However, the residual structure present in the pressure unfolded state of proteins (11, 12) could lead to a heterogeneous compressibility similar to that of the native state. The structural origin of the change in compressibility upon unfolding of proteins is still a topic of interest, and will be discussed further in this thesis. Additional studies will be required to measure precisely the small but perhaps significant contribution of compressibility to the pressure stability of proteins.

2. Physical origins of ΔV

The difference in volume between the folded and unfolded states, or ΔV, is the major determinant of the behavior of proteins under pressure. Because pressure and volume are conjugate variables, pressure unfolding equilibrium thermodynamics experiments at constant temperature yield information on the volume of the system according to equation (5). If the volume of the folded state is larger than that of the

unfolded state, then the ΔV, defined as $V_{unfolded} - V_{folded}$, will be negative. In this situation an increase in pressure will promote the unfolded state and lead to unfolding.

The unfolding profile of a protein, monitored with a spectroscopic probe such as Trp fluorescence or amide resonances in an HSQC spectrum measured a function of pressure, has a characteristic sigmoidal shape (Fig. 1). The slope of this transition is proportional to the ΔV of the protein; the more cooperative transitions report a higher ΔV value. The midpoint of the transition will be determined by both the ΔV and the global stability of the protein.

Typical ΔV values for proteins are negative under a variety of environmental conditions (43). The magnitude ranges from tens of mL mol^{-1} to about 200 mL mol^{-1}. This represents less than 1% of the total molar volume of a small protein near 15 kDa comparable to staphylococcal nuclease, the protein used for the present studies (44). A volume of 200 mL $mol^{-1}$is comparable to the molecular volume of 17 water molecules. Although small, this difference in volume between folded and unfolded states is sufficient to drive the unfolding of proteins.

The physical origins of ΔV that have been proposed are: (1) The loss of cavities that are present in the folded state and lost upon unfolding (13, 16, 45, 46); (2) the constriction of water around buried polar groups upon exposure to bulk water in the unfolded state, a phenomenon known as electrostriction (47); (3) the pressure dependent changes in the structure of bulk water that could lead to a weakened hydrogen bond network and thus a weakened hydrophobic effect (48); and (4) the possible changes in the expansivity (23, 24) and compressibility (25) of proteins that have been discussed above.

2.1 Cavities

The cores of protein are usually well packed (49, 50). In fact, the volume occupied by a residue in the protein interior is equivalent to that of its pure amino acid crystal (51). However, packing densities vary throughout the structure (turns, helices and strands: 0.794, 0.744, and 0.723, respectively) (52). This suggests that cavities and packing defects are distributed heterogeneously distributed over the protein structure. Some cavities are large enough to hold water molecules; nevertheless, in crystal structures they usually appear dry. This does not necessarily mean they are empty, but to the extent that they are devoid of water molecules, the disappearance of the cavities upon unfolding will contribute to a decrease in system volume.

2.1.1 Cavities in T4 lysozyme

In the late 80's and early 90's Brian Matthews and colleagues undertook the first systematic study of structural and thermodynamic properties of variants of T4 lysozyme with cavities of varying sizes (53, 54). The variants were obtained by single point mutations that truncate hydrophobic residues in core positions by substitution to Ala. The crystal structures of the variants are virtually identical to the parent structure, except for an enlargement of the volume of cavities and a few changes in the rotameric state of side

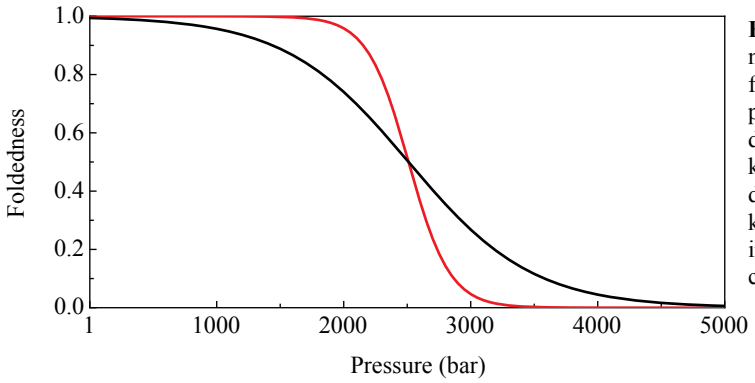


Figure 1. Simulated pressure unfolding of two model proteins. The y-axis represents a normalized folding parameter, with 1 representing a fully folded protein and 0 a fully unfolded one. The black curve describes unfolding of a protein with $\Delta G^{o} = 3$ kcal/mol and $\Delta V = 50$ mL/mol; the red curve describes the unfolding of a protein with $\Delta G^{o} = 9$ kcal/mol and $\Delta V = 150$ mL/mol. As pressure increases, both proteins unfold with different cooperativity, but the same transition midpoint.

chains lining the enlarged cavities. These studies demonstrated that in a stable globular protein it was possible to increase the volume of cavities without any significant effect on the global structure of the protein.

A subset of these variants was unfolded by pressure and monitored using Trp-fluorescence (45). The results were not always self-consistent and although general principles were published, they need to be corroborated in another protein. For example, a single substitution to Ala in the L99A variant quadruples the volume of the cavity observed by crystallography and results in a large experimental ΔV of 113 mL/mol. The structure of variant V149G shows a cavity comparable in size to that of L99A but the experimental ΔV is only 81 mL/mol. The authors argue that variant V149G reports a low ΔV owing to water molecues observed within the cavity. However, the structure of the L99G/E108V variant also shows that the cavity is filled with water molecules and yet it reports a much higher ΔV of 142 mL/mol. Whatever the role of cavity hydration might be, these data are not sufficient to account for the effect. Moreover, the A98L, with a small-to-large substitution intended to eliminate cavities, still reports a high ΔV value of 76 mL/mol. Unfortunately, the ΔV of the parent protein was never determined because the protein remained folded at the highest pressure accessible (~4 kbar). The problem is exacerbated by the fact that T4 lysozyme does not unfold in a two-state fashion, which impedes correlation of measured ΔV and ΔG^o (55). Despite the seemingly contradictory data yielded by this relatively small set of four variants, and despite the complexity inherent to the calculation of volumes from structure, the authors presented a linear relationship between experimental ΔV and the calculated volume of cavities (Fig. 2).

Studies by Font et al. using ribonuclease (RNase) A (56–58) and earlier by the Royer laboratory using only three variants of staphylococcal nuclease (SNase) (59) also yielded unclear results on the relationship between cavity volumes and the measured ΔV. In particular, variant V66L of SNase with a small-to-large substitution in the protein core intended to eliminate cavities reported a ΔV value of 96 mL/mol, almost 20 mL/mol greater than that of the parent protein (ΔV = 77 mL/mol).

2.1.2 Cavities in SNase

The first truly systematic study of the role of cavities in pressure unfolding involved ten variants of SNase engineered by substituting a core position to Ala (13). The crystal structures of all 10 variants were solved and used to confirm the increase in cavity volume with respect to the parent protein (Fig. 3). The ΔV values of all variants were measured by both Trp-fluorescence and NMR spectroscopy and they were all larger than that of the parent protein as expected.

This study with SNase demonstrated convincingly that cavities make a major contribution to the ΔV of unfolding. It also showed how the location of the artificial

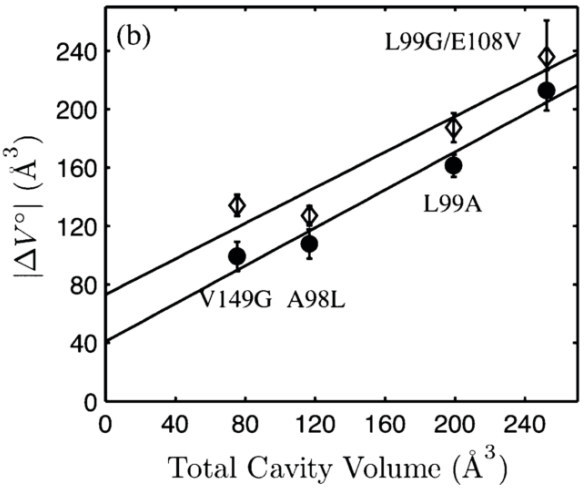


Figure 2: Correlation plot between the experimentally measured ΔV and the volume of cavities calculated from crystal structures of 4 variants of T4 lysozyme. The experimental ΔV on the y-axis was measured by pressure unfolding monitoring Trp fluorescence. The volume of cavities in crystal structures was obtained using the MSMS program (128) and a probe of 1.2 Å. The open diamonds and closed circles refer to experiments performed at 20 mM and 100 mM NaCl, respectively. The lines represents linear fits.

cavity influenced its contribution to ΔV – cavities closer to the protein-water interface seemed to make a lower contribution to ΔV. The site-specific information provided by high-pressure NMR spectroscopy also revealed a heterogeneous response to pressure and the ability of artificial cavities to modulate this cooperativity. The work established the importance of cavities in determining the ΔV of proteins and a need to better understand how the location, volume, polarity and state of hydration of cavities influence protein stability and pressure sensitivity.

However, extensive as this data set was compared to the Matthews data set, and regardless of how well behaved SNase is in equilibrium thermodynamic experiments of all sorts (60), the data set was still too sparse to allow establishing a relationship between the ΔV measured experimentally with the volume of engineered cavities or with the effects of the cavities on thermodynamic stability (Fig. 4).

2.2 Electrostriction

When an ionizable group becomes charged, the water around it constricts and becomes denser than bulk water (61). This phenomenon is known as electrostriction and results in a change in volume. Analogously, when a protein unfolds, the polar surface and ionizable residues that were interacting with protein become exposed to bulk water. As a result, the density of water around polar groups increases and yields a change in volume associated with the unfolding reaction. This contribution to the measured ΔV of proteins is thought to drive the pressure dissociation and perhaps even unfolding of proteins (19, 62–65).

The Wand laboratory used high pressure NMR spectroscopy to monitor the dissociation of a calmodulin-peptide complex under pressure (47). By following the signal of an interfacial Arg guanidinium group on the peptide, the authors measured a ΔV of ~60 mL/mol related to the dissociation of the complex (47, 66). They argued that the observed ΔV originated from the breaking of a putative salt bridge between an Arg side chain and the peptide at increased pressures, leading to an increase in hydration of the ionized residue in the dissociated state with a concomitant decrease in system volume. However, the observed change in ΔV could be due to eliminated cavity volume upon

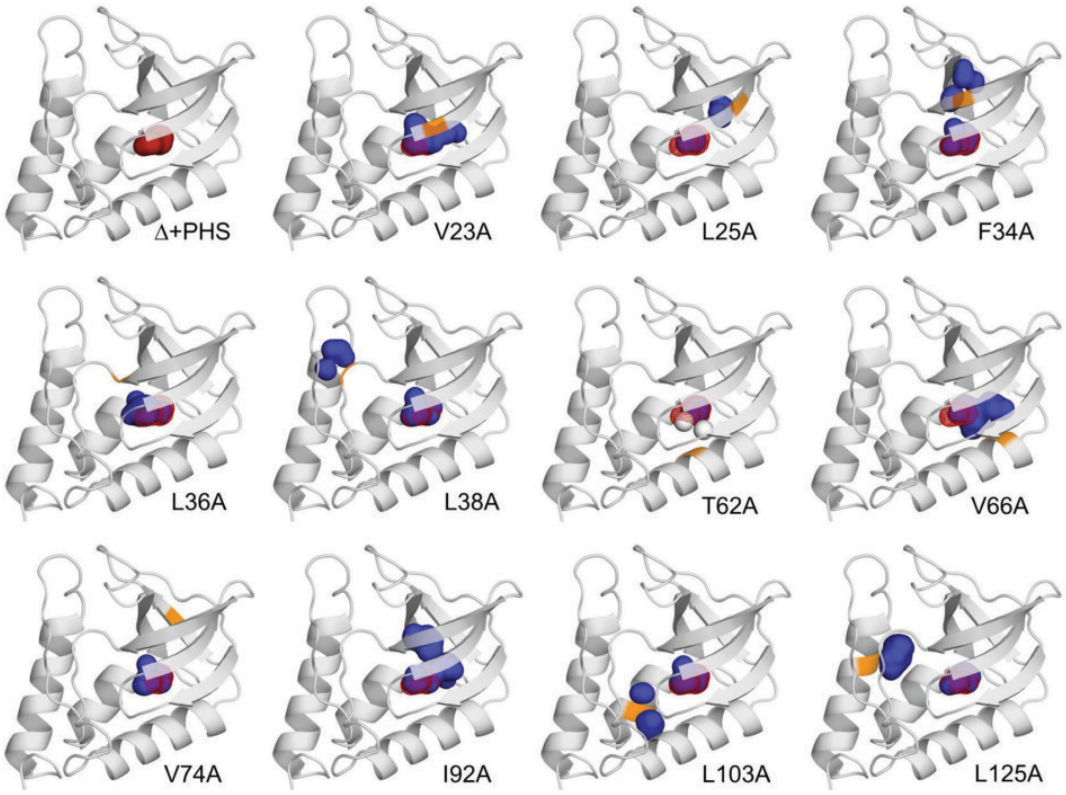


Figure 3. Crystal structures of SNase and its cavity-containing variants engineered by substitution of internal positions with Ala. The natural cavity is shown as red spheres. The cavities in the variants are shown with blue spheres. The location of the site of the Ala substitution is highlighted in orange. The T62A variant displays two internal water molecules, shown as white spheres. In some proteins the Ala substitution enlarges the natural cavity (e.g. I92A) but in others it generates a second cavity elsewhere in the protein (e.g. L125A).

exposure of the large Arg side chain and the concomitant change in exposed surface area at the protein-peptide interface. The fact remains that there are no clean experimental data with which to examine in quantitative detail the magnitude of contributions from electrostriction to the ΔV of unfolding.

Ionizable residues that are buried and neutral in the folded state of the protein and become charged in the unfolded state should contribute to solvation density differences. To estimate the contribution of such an ionization event to the ΔV of unfolding of a protein, variants of SNase with various buried ionizable and polar side chains were pressure unfolded (67). In the crystal structure of the V66D variant, the side chain of Asp-66 is buried in the hydrophobic core (68). Owing to the low polarity and polarizability of the protein interior, the pK_a of Asp-66 is 8.7 (69). Therefore, at pH 8, Asp-66 is mostly neutral in the folded state and changes ionization state upon unfolding. When the V66D variant was pressure unfolded at pH 8, the neutral Asp-66 became fully exposed to solvent, its pK_a became normalized (pK_a ~4.0), and the side chain became charged. The measured ΔV reported an increase of 20 mL/mol with respect to the parent protein, as one would expect from electrostriction. However, in this experiment, electrostriction was not the only possible contribution to ΔV. First, the experimental

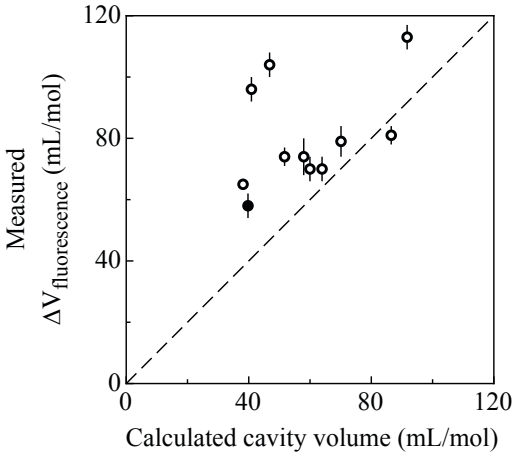


Figure 4: Correlation plot between the experimentally measured ΔV and the volume of cavities calculated from crystal structures of SNase and 10 variants with a substitution to Ala. The experimental ΔV on the y-axis was measured by pressure unfolding monitoring Trp fluorescence. The volume of cavities in crystal structures was obtained using the CASTp server and a probe of 1.4 Å. A line of slope 1 is shown as a dashed line. Although some of the points fall near the diagonal, no clear trend can be established even with this extensive set of 11 proteins.

conditions (pH, buffer and Gdn concentration) were not identical and could have contributed to the difference in measured ΔV. Second, the V66D substitution changes the van der Waals volume of the side chain and any increase in the volume of cavities would have resulted in an increase in ΔV. Third, and perhaps most perplexing, the variant V66N that bears an Asn at position 66, the neutral analog of Asp, was pressure unfolded and yielded an even greater increase in ΔV of 60 mL/mol with respect to SNase. One would expect that ionization of the buried Asp would result in a larger electrostrictive effect than simply the exposure of a neutral group. Instead, the data suggested that exposing a charge to bulk water upon unfolding of variant V66D decreased the ΔV of unfolding with respect to its neutral analog, V66N. The magnitude and even sign of the contribution to ΔV from the electrostriction of charges, whether buried or in a salt bridge, remains unclear.

Perhaps the main reason to consider changes in solvent density originating from hydration of exposed surfaces (i.e., electrostriction) as a significant contribution to the pressure stability of proteins is the dominant role that changes in solvent exposed surface area have in temperature and chemical denaturation of proteins. Transfer experiments of model compound amino acids to water from a hydrophobic medium like octanol that mimics the protein interior revealed that the difference in solvent exposed surface area of residues could be used to describe the heat capacity (ΔC_p) and the dependence of thermodynamic stability on concentration of chemical denaturant (*m*-value) of proteins (3, 70, 71). However, Baldwin and Kauzmann pointed out that the changes in volume associated with the transfer of hydrophobic molecules do not describe the behavior of proteins under pressure (2, 4, 72). Ordinary surface area calculations that prove capable

of reproducing the behavior of proteins under heat and chemical perturbation are simply not enough to describe the pressure thermodynamics of protein folding(6).

2.3 Properties of water under pressure

The phase diagram of water reflects structural changes above 10 kbar. It has been suggested that at pressures below 10 kbar, changes in the hydrogen bonding structure of water could lead to a reduced hydrophobic effect and thus to the unfolding of proteins under pressure (48, 73). However, the unfolding transitions observed by pressure denaturation can occur at a pressure as low as 500 bar (13, 74). Furthermore, the ΔV values obtained from techniques that can access pressures of 3 kbar are in agreement with those ΔV values obtained using pressure perturbation calorimetry (PPC), which involves pressure changes of only 5 bar (23, 75–77). Such low pressures are inconsistent with the structural changes needed in bulk water to account for the unfolding of proteins.

A systematic study of the pressure unfolding of a repeat protein, namely the ankyrin domain of the notch receptor, was performed to assess the contribution of each repeat to the protein's pressure sensitivity (78). The protein is composed of structural modules (repeats) with high sequence similarity. If the ΔV of the protein is due to a global property of bulk water, deletion of a repeat, regardless of where along the structure it is located, should yield the same effect on ΔV, and multiple repeat deletions should yield a cumulative effect. However, the data showed that the ΔV of the repeat protein was almost exclusively determined by the two central repeats, and that deletions from the N- or C-terminal had no measurable effect on the measured ΔV. This heterogeneous sensitivity to pressure of a repeat protein is inconsistent with a global change in a

property of water. Instead, the authors argued that the high ΔV of the central repeats was due to the large cavities they contained.

3. Cavities in proteins

The insights obtained from studies of the different structural and physical factors that determine the behavior of proteins under pressure have a common theme: the pressure response of proteins is highly location dependent. The location, volume and polarity of cavities appear repeatedly as the main structural determinant of this dependence. Although this thesis centers around the effects of pressure on proteins, the work advances the structural and dynamic studies of cavities and their effect on the local stability, conformational entropy and state of hydration of proteins.

3.1 Structural consequences of cavity formation

The first systematic study of the structural consequences of cavity formation was performed in the early 90's. Eriksson et al. substituted side chains in the core of T4 lysozyme by Ala and observed increases in the volume of cavities in the crystal structures of 25 variants (53, 54). The generated cavities resulted from Ala substitutions of Val, Ile, Leu, Met or Phe residues that were either fully solvent excluded or partially solvent exposed. The cavities generated by the Ala substitutions were located in both N- and C-terminal domains and resulted in only small rearrangements of the protein structure involving changes of rotameric states of side chains that tended to reduce the volume of the cavity.

The cavities were visualized in crystal structures and their volume calculated using a probe of radius 1.2 Å to define the probe-accessible surface. Some of the engineered cavities connected to the external surface of the protein and were accessible from bulk solvent. A few of them were even found hydrated by one or more water molecules. In these cases, the cavities were artificially sealed off from bulk using dummy atoms and volume calculations ignored the hydrating water molecules.

The thermodynamic stabilities of the variants were determined by thermal melts. Unfortunately these studies were undermined by the deviation from two-state behavior in the unfolding of T4 lysozyme (55, 79–81) in calorimetric measurements (82, 83), as well as by the inadequate extrapolation of the heat capacity over more than 40 °C (84). In spite of the ad-hoc approach, the cavity volumes obtained were correlated to the thermodynamic stabilities (ΔGs) and yielded an energetic cost of cavity generation of 37 cal/mL. This correlation included all cavities: hydrated cavities, pockets accessible by bulk solvent, cavities in independent domains of a protein, and cavities lined by polar groups. This pioneering study made important contributions, but it is clear that this need to be followed by a more detailed study of the consequences of cavity formation on thermodynamic stability and on structure. That is one of the goals of the experiments discussed in this dissertation.

3.2 Dynamic consequences of cavity formation: crystallography

Owing to the high packing efficiency of proteins it is difficult to improve stability by increasing packing density (50). Overpacking with substitutions of small side chains with larger ones generally leads to a decrease in packing efficiency, as evidenced by the

volume of cavities observed by crystallography, decrease in thermodynamic stability, and increase in measured ΔV values (45, 85–88).

Overpacking also leads to an increase in motion of core side chains (89–91). The Fox laboratory engineered a variant of SNase with a substitution to Cys at the internal position 23 and subsequently modified the side chain to attach longer groups via a disulfide bond. The prosthetic groups increased in size one methylene at a time from methane up to pentane. The crystal structures showed that the electron density for the prosthetic group became less well-defined with increasing side chain length, along with localized distortions of the backbone and rotameric changes of neighboring side chains while the rest of the protein structure remained unchanged. The stability of the methyl reacted variant showed an increase of 1.8 kcal/mol, suggestive of an improvement in packing of the core. However, the subsequent variants with increasing length of the side chain at position 23 showed a concomitant decrease in stability, with the propane reacted variant showing similar stability to the V23C background and the pentane reacted variant showing the lowest stability with a further decrease of 1.3 kcal/mol. Some of the T4 lysozyme cavity-filling variants studied by the Matthews laboratory (87) also showed evidence of disorder at the substitution site with a concomitant decrease in global stability, such as variant L99I. These early studies began to elucidate how the volume of cavities influences the conformational entropy of buried side chains.

In order to quantitatively assess the disorder observed in the side chains of crystal structures, the Alber laboratory developed an algorithm to inspect more closely the electron density obtained for crystal structures of 30 different proteins (92–94). They found that in structures from diffraction data collected at room temperature, side chains

often display alternative conformations, and these motions involved more rotamers and often completely different rotamers from those observed in cryogenic crystals. Buried side chains were also affected, with longer side chains showing increased occupancy of alternative rotamers as compared to the cryogenic structure. Cryocooling appears to induce an increase in packing efficiency of the protein and a reduction in cavity volume (94). In contrast with the overpacking experiments by Fox and Matthews, elimination of cavities through cryocooling appears to reduce the dynamics of buried side chains with respect to room temperature. This could be analogous to the compression of cavities and the restriction of fluctuations that give rise to cavities at high pressure.

3.3 Dynamic consequences of cavity formation: NMR spectroscopy

The recent commercialization of equipment for high pressure solution NMR spectroscopy has now enabled routine studies of pressure effects on proteins, but until recently these have been few and far between. NMR spectroscopy and the relaxation parameters that describe motions in the nanosecond timescale offer another direct measure of the dynamics in the interior of proteins and their pressure sensitivity. Deuterated methyls on the side chains of Ala, Val, Thr, Ile, Leu and Met can be used to track the range of motions experienced by these side chains as described by an order parameter that describes the modes of side chain motions (95–102). Clustering of side chains with similar dynamics would indicate a coupling of protein motions through entropic energy and has been proposed as the allosteric mechanism of enzymes (103, 104).

In principle, pressure will reduce the local fluctuations that result in the transient formation of cavities and favor the population of alternative states of lower volume. The Wand laboratory monitored the methyl dynamics of ubiquitin (Ub) at 2.5 kbar and observed a slight depletion of highly mobile side chains (36). As pressure was increased, 25% of the methyl bearing side chains showed a linear increase in motion, while 52% showed a decrease in motion. The protein core was enriched in the former, suggesting that specific regions in the core became more dynamic as a function of pressure, perhaps reflecting the heterogeneous distribution of packing defects. Interestingly, 23% of the methyl bearing side chains showed a non-linear response to pressure, with an initial increase before decreasing, much like what is observed with temperature (101). This non-linear response in the dynamics could reflect localized transitions or population of low-lying excited states that form part of the compression mechanism of proteins.

The Akasaka group studied the variant L99A of T4 lysozyme to understand the effects of engineering a cavity on the conformational fluctuations of the protein (105). By monitoring the maximum peak intensity of ^{1}H-^{13}C HSQC spectra as a function of pressure they observed a fast decrease in intensity of methyls highly localized around the cavity. This is consistent with their previous findings of large, non-linear chemical shifts of amide proton HSQC resonances corresponding to sites located near cavities in hen egg white lysozyme(34, 106). The authors argued that cavities enable specific conformational fluctuations that are necessary to populate a low-lying excited state previously identified by the Kay group (107). However, an alternative view on the effect of pressure on variant L99A of T4 lysozyme comes from the Wand laboratory (16), where they argued that the engineered cavity induces partial unfolding with pressure that exposes the cavity to bulk

water. Further studies in this vein will clarify how the local conformational fluctuations that result in transient cavities or access to internal cavities help determine the pressure behavior of proteins.

4. Hydration state of cavities studied with pressure

One of the intriguing question regarding internal cavities in proteins is their state of hydration. If internal cavities are hydrated their contributions to the ΔV of unfolding will be minimized. It has also been proposed that pressure drives the penetration of water molecules into the protein core and disrupts the favorable hydrophobic interactions that keep the protein folded (108). Pressure can indeed be used to fill cavities with molecules of noble gases (109). Low pressures of only a few dozen bar were required to observe the localization of xenon and argon in the cavities of wild type T4 lysozyme and in six of its cavity-containing variants. Neon, however, which is comparable in size to a water molecule, cannot be observed in the artificial cavities of T4 lysozyme variants even at the highest pressures accessed (64 bar). In the case of water, the absence of hydrogen bonds in the hydrophobic cavity should diminish the ability of water to localize within non-polar cavities; even the placement of a single water molecule in a completely hydrophobic environment is extremely unfavorable (110, 111).

4.1 Cavity hydration studies with high pressure crystallography

The Matthews and Gruner laboratories (15) used high-pressure (2000 bar) crystallography to study the hydration state of the engineered cavity in the L99A variant of T4 lysozyme. They subjected crystals of the protein to high pressure and monitored

how the electron density within the cavity changed. Structurally, the artificial cavity showed no change in the volume or shape. The authors observed an increase in electron density within the cavity, which they ascribed to up to 3 new, buried water molecules. However, the poorly defined density yielded modeled water molecules with very high B-factors (4-5 fold that of the Cβ of Ala-99), making these waters highly dubious. It is difficult to ascertain the origin of delocalized electron density in a crystallographic experiment, since cosolvent such as β-mercaptoethanol, used in the mother liquor of this experiment to precipitate and induce crystallization, can readily localize within internal cavities. It is possible now to download the structure factors and coordinates from RCSB and model in β-mercaptoethanol in place of the dubious waters. The cosolvent fits well in the electron density found in the cavity and a single round of refinement results in an improvement of the map and the statistics, making β-mercaptoethanol the clear choice to model over water.

4.2 Cavity hydration studies with high pressure NMR spectroscopy

To investigate the state of hydration of ubiquitin, the Wand laboratory confined ubiquitin (Ub) to a reverse micelle, which allows direct detection of protein-water interactions through the nuclear Overhauser effect (112, 113). Interaction between water and surface amide and carbon protons span a range of dynamics, from long-lived interactions with structural waters observed by crystallography to interactions too dynamic to detect (the detection limit of this technique requires a residence time of about 100 ps for the water-protein NOE interaction). Most amide protons in the protein core did not show detectable interactions with solvent, indicating that if any buried waters were

present, they displayed dynamics faster than the detectable picosecond timescale. A highly dynamic core is consistent with some hydrophobic side chains in the core showing low order parameters (36).

The Wand laboratory also used high pressure NMR spectroscopy to study the response of the cavity-containing variant L99A of T4 lysozyme, confined to a reverse micelle with about 4 layers of water surrounding it (16). When the variant L99A was confined, its pressure unfolding was suppressed because the radius of gyration of the unfolded state was larger than that of the micelle (~32 Å vs. 25 Å). This is not unlike the consequences of confining a protein to a crystal lattice, as studied by Matthews and Gruner (15) and Nagae et al. (35). But unlike the high pressure crystallographic studies, Nucci et al. found clear evidence for localized water molecules within the artificial cavity of variant L99A of T4 lysozyme. To observe buried waters, however, the authors had to increase the pressure to values above the unfolding transition of the protein when free in solution. Furthermore, the authors clarify that it is not possible to distinguish whether the observation is of pressure induced water penetration or of dynamic restriction of internal waters. Thus, cavities can be hydrated, but at least in the case of this protein this appears to occur only at pressures higher than those needed to unfold the protein when free in solution. Alternatively, the effect of pressure might be to restrict the dynamics of internal water molecules, which might already be present at 1 bar.

4.3 Cavity hydration studies with simulations at high pressure

Molecular dynamics simulations have proven useful to investigate the mechanisms of hydration of proteins (111, 114–119). In particular, work by the Halle

group (119) showed how millisecond long simulations, one of the longest simulations to date, could identify water wires and transient tunnels through bovine pancreatic trypsin inhibitor. This is consistent with simulations of SNase variants with buried polar and ionizable groups, which show that preferred pathways exist through which waters from bulk access the buried polar group and that such routes are enabled by localized structural fluctuations that result in the formation of transient cavities (117).

NMR spectroscopy was used to solve the solution structure of Ub at pressures below the unfolding transition. The structure underwent a slight perturbation at 3000 bar relative to the structure at 30 bar, with RMSD values of up to 3 Å. At 3000 bar the helix composed of residues 23-34 shifts away from the protein core and partially exposes regions of the protein previously buried (42). Ensembles of NMR structures of Ub determined at 1 and 3000 bar were used by Day & García in simulations under pressures of 1, 3000 and 6000 bar to monitor the change in hydration of the protein core (120). Simulations at 1 bar of the structures determined at 1 bar and at 3000 bar show an average of 20 water molecules penetrating the protein core. However, when the pressure at which the simulation is run is increased, first to 3000 bar and then to 6000 bar, the average number of waters penetrating the structures solved at 3000 bar increases to 24 and to 25, respectively, while the number of penetrating waters in the structures solved at 1 bar remains constant up to 3000 bar and increases only to 21 at 6000 bar. Concomitant with a higher number of penetrating water, the volume of the high pressure ensemble decreases faster with increasing pressure than that of the 1 bar ensemble. Interestingly, this study found that the volume of water molecules increased upon transfer from bulk water into the protein interior, and this change in volume increased as a function of

pressure to up to 1.1 mL/mol. Albeit small, this volume change already argues against a pressure driven hydration of the protein core (108).

5. Considerations on the effect of pressure on the unfolded state of proteins

The thermodynamics of protein folding relies as much on the properties of the folded state as on those of the unfolded state. Experimental difficulties have limited the information available on the unfolded state of proteins and have led to a native-centric view of protein folding (121). The effect of pressure on the behavior of proteins could stem from changes induced in the unfolded state. If the properties such as volume, compressibility or expansivity remain largely unchanged as a function of pressure in a well-ordered, structured and folded native state, then the pressure sensitivity must stem from changes in the packing or solvation of the unfolded state. Furthermore, the residual structure observed for the pressure unfolded state could be affected by changes in the amino acid sequence of the protein (11). Small-angle X-ray scattering and infrared spectroscopy studies on SNase indicate that this residual structure appears to be insensitive to pressures as high as 8000 bar (122). They also find that the cavity-generating substitution V66A that leads to a significant change in ΔV (13) does not appear to affect the absolute values of the infrared signatures of the unfolded state, nor does it appear to affect their dependence on pressure (122). Substitutions of larger or multiple internal side chains with Ala, substitutions to bulkier side chains, or substitutions of internal hydrophobic side chains with polar residues could, in principle, leave the partial molar volume of the native state unchanged but alter the volume of the unfolded state. This view of protein unfolding is consistent with the work by Shortle on variants of

SNase with substitutions that alter the dependence of stability on chemical denaturant concentration, *m* (71). Although outside the scope of this dissertation, studies of the effect of amino acid substitutions on the volume of the unfolded state are necessary for a comprehensive understanding of pressure effects in proteins.

6. Previous pressure studies with staphylococcal nuclease

Much of what is known about protein folding has come as a direct result of understanding the structural origins of the effect of temperature and chemical denaturants on proteins. A good example of this is the possibility to perform energy calculations of protein ensembles from structure (123). However, the physical and structural factors that govern temperature or chemical denaturation are not the same as those that govern pressure sensitivity of proteins (6, 7). Pressure and temperature are fundamental parameters and are equally important for the description of the thermodynamics of a system. The lack of understanding of the physical origins of pressure effects in proteins limits the interpretation of volumetric data and the information it holds on hydration and conformational fluctuations of proteins.

Staphylococcal nuclease (SNase) has served previously as the protein of choice to examine the correlation between changes in surface area upon unfolding and ΔV. Variants of SNase with internal ionizable groups were also used previous in attempts to determine the contributions from electrostriction to ΔV (67). The only direct measurement of the change in compressibility upon unfolding using densitometry (25) was performed using SNase, as well as the first study of the effect of artificial cavities on expansivity (24). SNase was also used for a systematic study of the relationship between

the calculated volume of cavities observed in structures and the ΔV measured experimentally (13). Having demonstrated its utility as a model system for detailed examination of determinants of pressure effects, we selected SNase for the studies outlined in this dissertation.

7. Structural origins of pressure effects in proteins: structure of this dissertation

A large volume of work has established beyond reasonable doubt that the internal cavities that are present in the folded state and are eliminated upon unfolding are one of the main determinants of the pressure sensitivity of proteins. On the other hand, even the relatively high throughput studies by Matthews and coworkers (13, 45, 56, 58, 124) did not lead to generalizable principles with which to interpret the structural basis of pressure sensitivity of proteins. The research described in this dissertation examines in unprecedented detail the relationship between presence of cavities in the folded state and pressure sensitivity.

This study is divided into three sections. Chapter 2 describes a quantitative relationship between the experimentally measured ΔV and the volume of cavities observed in crystal structures obtained at 1 atm under cryogenic conditions. This involved a systematic study of 20 variants of SNase in which 1, 2 or 3 internal positions were substituted with Ala to generate artificial cavities of increasing volume. Crystal structures were solved to determine the volume of internal cavities, their location and their general properties. Pressure unfolding monitored with Trp fluorescence or with ^{1}H-^{15}N HSQC spectra were performed to determine ΔV of unfolding for all variants. This is an extension of a previous study focused on the 10 variants with single Ala substitution

(13). That initial study was revealing as it showed a clear increase in ΔV of all the variants with respect to the parent protein. However, the results of variants with a single Ala substitution did not allow determination of a clear correlation between the measured ΔV values and the volume of cavities observed in the crystal structures (Fig 4). With the larger artificial cavities generated by multiple Ala substitutions resulting in a three-fold increase in ΔV with respect to the parent protein, it was possible to establish a linear trend between measured ΔV and the volume of cavities calculated from structures.

Chapter 3 reports on attempts to eliminate ΔV by filling the naturally occurring cavity using multiple small-to-large side chain substitutions. Variants were engineered with bulky side chains such as Phe and Met substituting at core positions where the bulky side chains would point towards the natural microcavity of the protein. Crystal structures showed that the side chains indeed eliminate the microcavity beyond detection using standard algorithms. Pressure unfolding experiments reported decreases in ΔV of nearly 30 mL/mol. In one of the variants the ΔV was too small to determine accurately. The variants extended the linear relationship between measured and calculated ΔV reported in Chapter 2. The data strongly suggest that the ΔV of SNase is determined primarily by the volume of internal cavities, without the need to invoke contributions of any other nature.

Chapter 4 presents results from a study to evaluate the role of internal hydration as a factor that affects the volume of internal cavities and therefore the pressure sensitivity of proteins. Substitutions with polar residues were used to line cavities with polar residues that promote the internal state of hydration of the cavities. The ΔV of a variant with a cavity filled with water molecules was expected to be smaller than the ΔV of variants in which the more hydrophobic cavities are empty. In contrast, there could be

an increase in ΔV of the polar cavity variant due to the electrostriction of the buried polar side chains and buried water molecules upon unfolding and release into bulk water. If in the folded state these engineered polar residues cannot satisfy their hydrogen bonds as well as in the unfolded state, unfolding and release of the buried polar residues into bulk water will result in improved hydration and an accompanying decrease in volume.

The crystal structures of these variants showed how buried polar groups led to various patterns of internal hydration. The results showed that the pressure unfolding of variants showing increased internal hydration did not report a smaller ΔV. The data suggested instead that water molecules are present in cavities even in the absence of the polar substitutions, but are too dynamic to detect by crystallography.

Lastly, chapter 5 summarizes the materials and methods used throughout the thesis work and the Appendix 1 explores the role of cavities as determinants of the conformational state and pK_a of internal ionizable residues.

Each chapter in this thesis addresses a fundamental question about determinants of pressure sensitivity of staphylococcal nuclease. The experimental data from this study is essential to begin to exploit the volumetric information available through pressure perturbation that describes the hydration and conformational fluctuations of proteins, and to complement the protein phase diagram and the underlying knowledge of protein folding.

The work described in this dissertation represents an extension of on-going studies we have performed in collaboration with the Royer laboratory at the Centre de Biochimie Structurale, Montpellier, France (presently at Rensselaer Polytechnic Institute, Troy, NY). I am only including unpublished data in this dissertation. Results from

previous studies have been presented elsewhere. In the following list of publications, my work involved the engineering and purification of the protein variants, the determination of the crystal structures, the measurement of protein stability by GdnHCl unfolding experiments, and participation in the data analysis and the making of figures:

(1) Julien Roche, **José A. Caro**, Douglas R. Norberto, Philippe Barthe, Christian Roumestand, Jamie L. Schlessman, Angel E. García, Bertrand García-Moreno E., and Catherine A. Royer. (2012) Cavities determine the pressure unfolding of proteins. *Proc Natl Acad Sci USA* 109:6945–50.

This paper describes a systematic study of the relationship between the calculated volume of cavities observed in structures and the ΔV measured experimentally with pressure unfolding monitored with Trp-fluorescence or with ^{1}H-^{15}N HSQC spectra. The study illustrates the magnitude of ΔV associated with single amino acid substitutions in the hydrophobic interior of a protein.

(2) Julien Roche, **José A. Caro**, Mariano Dellarole, Ewelina Guca, Catherine A. Royer, Bertrand García-Moreno E., Angel E. García, and Christian Roumestand. (2012) Structural, energetic, and dynamic responses of the native state ensemble of staphylococcal nuclease to cavity-creating mutations. *Proteins* 135:14610–14618.

High pressure NMR spectroscopy was used in this study to investigate the structural responses to pressure in 4 variants of SNase with artificial cavities (125). The response of the protein to pressure depended strongly on the location of the cavity, which could be explained by the destabilization of cooperative subunits and the local conformational excursions characterized by MD simulations.

(3) Julien Roche, Mariano Dellarole, **José A. Caro**, Ewelina Guca, Douglas R. Norberto, Yinshan Yang, Angel E. García, Christian Roumestand, Bertrand García-Moreno, and Catherine A. Royer. (2012) Remodeling of the Folding Free Energy Landscape of Staphylococcal Nuclease by Cavity-Creating Mutations. *Biochemistry* 51:9535–9546.

This study further studied the cooperativity of the pressure unfolding of SNase and its cavity containing variants using hydrogen exchange NMR spectroscopy at a range of pressures (126). The protection factors as a function of pressure can indicate regions of increased local fluctuations that result in the population of partially folded states. Variant L125A, which displays an artificial cavity near the surface of the protein, the cavity increased the pressure sensitivity of the protein in a highly localized fashion. Cavities deep in the core of SNase, such as the one in variant I92A, appeared to increase exchange across the entire protein, suggesting that a high number of low-lying excited states were enabled by the substitution to Ala.

(4) Julien Roche, Mariano Dellarole, **José A. Caro**, Douglas R. Norberto, Angel E. García, Bertrand García-Moreno E., Christian Roumestand, and Catherine A. Royer. (2013) Effect of Internal Cavities on Folding Rates and Routes Revealed by Real-time Pressure-Jump NMR Spectroscopy. *J Am Chem Soc* 82:1069–1080.

In this study the site-specific folding kinetics of SNase were characterized using high pressure NMR spectroscopy. The artificial cavities were shown to exert a strong effect on the volume of the transition state ensemble, indicating that different folding routes must exist (127).

(5) Mariano Dellarole, Kei Kobayashi, Jean-Baptiste Rouget, **José A. Caro**, Julien Roche, Mohammad M. Islam, Bertrand García-Moreno E., Yutaka Kuroda, and

Catherine A. Royer. (2013) Probing the Physical Determinants of Thermal Expansion of Folded Proteins. *J Phys Chem B*.

This paper reports on the expansivity of variants of SNase with artificial cavities, studied by PPC. The work showed that a single cavity generating substitution in SNase resulted in a much larger change in expansivity than the change observed for 11 surface substitutions to Ala in bovine pancreatic trypsin inhibitor (BPTI). This indicated that the expansivity of folded proteins is highly sensitive to the tertiary structure of the protein (24). It also showed that SNase variants I92A and L125A, which report dramatically different ΔV values, reported identical PPC traces, indicating that expansivity is not correlated with ΔV.

(6) Mariano Delarolle, **José A. Caro**, Julien Roche, Martin Fossat, Philippe Barthe, Catherine A. Royer, and Christian Roumestand. (2014) Protein Thermal Expansion Reveals the Architecture of Intramolecular Interactions. *Submitted to JACS.*

The determinants of the expansivity of cavity-containing variants of SNase were further explored using high pressure NMR spectroscopy. To this end the variants of SNase with artificial cavities were pressure unfolded at various temperatures. Increases in temperature resulted in decreases of the average ΔV values, and SNase variant I92A/L125A, with the largest artificial cavities, reported the largest temperature dependence. This could be explained by the decrease in the strength of hydrogen bonds due to increased exposure by artificial cavities. As a consequence, a less constrained native structure would undergo greater thermal expansion.

Bibliography

1. Bridgman PW (1914) The coagulation of albumen by pressure. *J. Biol. Chem.* 19(4):511–512.

2. Brandts JF, Oliveira RJ, Westort C (1970) Thermodynamics of Protein Denaturation. Effect of Pressure on the Denaturation of Ribonuclease A". *Biochemistry* 9(4):1038–1047.

3. Kauzmann W (1959) Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* 14:1–63.

4. Zipp A, Kauzmann W (1973) Pressure Denaturation of Metmyoglobin. *Biochemistry* 12(21):4217–4228.

5. Frye KJ, Perman CS, Royer CA (1996) Testing the correlation between ΔA and ΔV of protein unfolding using m value mutants of staphylococcal nuclease. *Biochemistry* 35(31):10234–10239.

6. Kauzmann W (1987) Thermodynamics of Unfolding. *Nature* 325:763–764.

7. Baldwin RL (1986) Temperature dependence of the hydrophobic interaction in protein folding. *Proc. Natl. Acad. Sci. U.S.A.* 83(21):8069–8072.

8. Foguel D, Suarez MC, Ferrão-Gonzales AD, Porto TCR, Palmieri L, Einsiedler CM, Andrade LR, Lashuel H a, Lansbury PT, Kelly JW, Silva JL (2003) Dissociation of amyloid fibrils of alpha-synuclein and transthyretin by pressure reveals their reversible nature and the formation of water-excluded cavities. *Proc. Natl. Acad. Sci. U.S.A.* 100(17):9831–9836.

9. Niraula TN, Konno T, Li H, Yamada H, Akasaka K, Tachibana H (2004) Pressure-dissociable reversible assembly of intrinsically denatured lysozyme is a precursor for amyloid fibrils. *Proc. Natl. Acad. Sci. U.S.A.* 101(12):4089–4093.

10. Ohmae E, Murakami C, Tate S, Gekko K, Hata K, Akasaka K, Kato C (2012) Pressure dependence of activity and stability of dihydrofolate reductases of the deep-sea bacterium Moritella profunda and Escherichia coli. *Biochim. Biophys. Acta (BBA) - Proteins Proteomics* 1824(3):511–519.

11. Zhang J, Peng X, Jonas A, Jonas J (1995) NMR study of the cold, heat, and pressure unfolding of ribonuclease A. *Biochemistry* 34(27):8631–8641.

12. Vajpai N, Nisius L, Wiktor M, Grzesiek S (2013) High-pressure NMR reveals close similarity between cold and alcohol protein denaturation in ubiquitin. *Proc. Natl. Acad. Sci. U.S.A.* 110(5):E368–E376.

13. Roche J, Caro JA, Norberto DR, Barthe P, Roumestand C, Schlessman JL, Garcia AE, García-Moreno E. B, Royer CA (2012) Cavities determine the pressure unfolding of proteins. *Proc. Natl. Acad. Sci. U.S.A.* 109(18):6945–6950.

14. Babu CR, Hilser VJ, Wand AJ (2004) Direct access to the cooperative substructure of proteins and the protein ensemble via cold denaturation. *Nat. Struct. Mol. Biol.* 11(4):352–357.

15. Collins MD, Hummer G, Quillin ML, Matthews BW, Gruner SM (2005) Cooperative water filling of a nonpolar protein cavity observed by high-pressure crystallography and simulation. *Proc. Natl. Acad. Sci. U.S.A.* 102(46):16668–16671.

16. Nucci NV, Fuglestad B, Athanasoula EA, Wand AJ (2014) Role of cavities and hydration in the pressure unfolding of T4 lysozyme. *Proc. Natl. Acad. Sci. U.S.A.* 111(38):13846–13851.

17. Hawley SA (1971) Reversible pressure-temperature denaturation of chymotrypsinogen. *Biochemistry* 10(13):2436–2442.

18. Smeller L (2002) Pressure-temperature phase diagrams of biomolecules. *Biochim. Biophys. Acta (BBA) - Protein Struct. Mol. Enzymol.* 1595(1-2):11–29.

19. Chalikian TV, Macgregor Jr RB (2009) Origins of pressure-induced protein transitions. *J. Mol. Biol.* 394(5):834–842.

20. Lin L-N, Brandts JF, Brandts JM, Plotnikov V (2002) Determination of the volumetric properties of proteins and other solutes using pressure perturbation calorimetry. *Anal. Biochem.* 302(1):144–160.

21. Mitra L, Smolin N, Ravindra R, Royer CA, Winter R (2006) Pressure perturbation calorimetric studies of the solvation properties and the thermal unfolding of proteins in solution-experiments and theoretical interpretation. *Phys. Chem. Chem. Phys.* 8(11):1249–1265.

22. Barrett DG, Minder CM, Mian MU, Whittington SJ, Cooper WJ, Fuchs KM, Tripathy A, Waters ML, Creamer TP, Pielak GJ (2006) Pressure perturbation calorimetry of helical peptides. *Proteins Struct. Funct. Bioinforma.* 63(2):322–326.

23. Vasilchuk D, Pandharipande PP, Suladze S, Sanchez-Ruiz JM, Makhatadze GI (2014) Molecular determinants of expansivity of native globular proteins: a pressure perturbation calorimetry study. *J. Phys. Chem. B* 118(23):6117–6122.

24. Dellarole M, Kobayashi K, Rouget J-B, Caro JA, Roche J, Islam MM, García-Moreno E. B, Kuroda Y, Royer CA (2013) Probing the Physical Determinants of Thermal Expansion of Folded Proteins. *J. Phys. Chem. B* 117(42):12742–12749.

25. Seemann H, Winter R, Royer CA (2001) Volume, expansivity and isothermal compressibility changes associated with temperature and pressure unfolding of Staphylococcal nuclease. *J. Mol. Biol.* 307(4):1091–1102.

26. Prehoda KE, Mooberry ES, Markley JL (1998) Pressure denaturation of proteins: evaluation of compressibility effects. *Biochemistry* 37(17):5785–5790.

27. Akasaka K, Li H, Yamada H, Li R, Thoresen T, Woodward CK (1999) Pressure response of protein backbone structure. Pressure-induced amide 15N chemical shifts in BPTI. *Protein Sci.* 8(10):1946–53.

28. Akasaka K, Li H (2001) Low-Lying Excited States of Proteins Revealed from Nonlinear Pressure Shifts in 1H and 15N NMR. *Biochemistry* 40(30):8665–8671.

29. Lassalle MW, Yamada H, Akasaka K (2000) The pressure-temperature free energy-landscape of staphylococcal nuclease monitored by (1)H NMR. *J. Mol. Biol.* 298(2):293–302.

30. Heremans K, Smeller L (1998) Protein structure and dynamics at high pressure. *Biochim. Biophys. Acta* 1386(2):353–70.

31. Akasaka K (2003) Highly Fluctuating Protein Structures Revealed by Variable-Pressure Nuclear Magnetic Resonance. *Biochemistry* 42(37):10875–10885.

32. Linowski JW, Liu N, Jonas J (1976) Pressure dependence of the proton NMR chemical shift in liquid water. *J. Chem. Phys.* 65(8):3383–3384.

33. Boonyaratanakornkit BB, Park CB, Clark DS (2002) Pressure effects on intra- and intermolecular interactions within proteins. *Biochim. Biophys. Acta (BBA) - Protein Struct. Mol. Enzymol.* 1595(1):235–249.

34. Kamatari YO, Smith LJ, Dobson CM, Akasaka K (2011) Cavity hydration as a gateway to unfolding: an NMR study of hen lysozyme at high pressure and low temperature. *Biophys. Chem.* 156(1):24–30.

35. Nagae T, Kawamura T, Chavas LMG, Niwa K, Hasegawa M, Kato C, Watanabe N (2012) High-pressure-induced water penetration into 3-isopropylmalate dehydrogenase. *Acta Crystallogr. Sect. D - Biol. Crystallogr.* 68(3):300–309.

36. Fu Y, Kasinath V, Moorman VR, Nucci NV, Hilser VJ, Wand AJ (2012) Coupled motion in proteins revealed by pressure perturbation. *J. Am. Chem. Soc.* 134(20):8543–8550.

37. Collins MD, Quillin ML, Hummer G, Matthews BW, Gruner SM (2007) Structural rigidity of a large cavity-containing protein revealed by high-pressure crystallography. *J. Mol. Biol.* 367(3):752–763.

38. Kundrot CE, Richards FM (1987) Crystal structure of hen egg-white lysozyme at a hydrostatic pressure of 1000 atmospheres. *J. Mol. Biol.* 193(1):157–170.

39. Samarasinghe SD, Campbell DM, Jonas A, Jonas J (1992) High-resolution NMR study of the pressure-induced unfolding of lysozyme. *Biochemistry* 31(34):7773–8.

40. Fuentes EJ, Wand AJ (1998) Local stability and dynamics of apocytochrome b562 examined by the dependence of hydrogen exchange on hydrostatic pressure. *Biochemistry* 37(28):9877–9883.

41. Fuentes EJ, Wand AJ (1998) Local Dynamics and Stability of Apocytochrome b562 Examined by Hydrogen Exchange. *Biochemistry* 2960(97):3687–3698.

42. Kitahara R, Yokoyama S, Akasaka K (2005) NMR snapshots of a fluctuating protein structure: ubiquitin at 30 bar-3 kbar. *J. Mol. Biol.* 347(2):277–285.

43. Royer CA (2002) Revisiting volume changes in pressure-induced protein unfolding. *Biochim. Biophys. Acta (BBA) - Protein Struct. Mol. Enzymol.* 1595(1):201–209.

44. Filfil R, Chalikian TV (2000) Volumetric and spectroscopic characterizations of the native and acid-induced denatured states of staphylococcal nuclease. *J. Mol. Biol.* 299(3):827–842.

45. Ando N, Barstow B, Baase WA, Fields A, Matthews BW, Gruner SM (2008) Structural and thermodynamic characterization of T4 lysozyme mutants and the contribution of internal cavities to pressure denaturation. *Biochemistry* 47(42):11097–11109.

46. Lassalle MW, Yamada H, Morii H, Ogata K, Sarai a, Akasaka K (2001) Filling a cavity dramatically increases pressure stability of the c-Myb R2 subdomain. *Proteins Struct. Funct. Bioinforma.* 45(1):96–101.

47. Urbauer JL, Ehrhardt MR, Bieber RJ, Flynn PF, Wand AJ (1996) High-Resolution Triple-Resonance NMR Spectroscopy of a Novel Calmodulin-Peptide Complex at Kilobar Pressures. *J. Am. Chem. Soc.* 118(45):11329–11330.

48. Grigera JR, McCarthy AN (2010) The behavior of the hydrophobic effect under pressure and protein denaturation. *Biophys. J.* 98(8):1626–1631.

49. Richards FM (1977) Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* 6:151–176.

50. Richards FM (1974) The interpretation of protein structures: total volume, group volume distributions and packing density. *J. Mol. Biol.* 82(1):1–14.

51. Chothia C (1975) Structural invariants in protein folding. *Nature* 254(5498):304–308.

52. Fleming PJ, Richards FM (2000) Protein packing: dependence on protein size, secondary structure and amino acid composition. *J. Mol. Biol.* 299(2):487–498.

53. Xu J, Baase WA, Baldwin E, Matthews BW (1998) The response of T4 lysozyme to large-to-small substitutions within the core and its relation to the hydrophobic effect. *Protein Sci.* 7(1):158–177.

54. Eriksson AE, Baase WA, Zhang XJ, Heinz DW, Blaber M, Baldwin EP, Matthews BW (1992) Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* 255(5041):178–183.

55. Llinás M, Gillespie B, Dahlquist FW, Marqusee S (1999) The energetics of T4 lysozyme reveal a hierarchy of conformations. *Nat. Struct. Biol.* 6(11):1072–1078.

56. Font J, Benito A, Torrent J, Lange R, Ribó M, Vilanova M (2006) Pressure- and temperature-induced unfolding studies: thermodynamics of core hydrophobicity and packing of ribonuclease A. *Biol. Chem.* 387(3):285–96.

57. Kurpiewska K, Font J, Ribó M, Vilanova M, Lewiński K (2009) X-ray crystallographic studies of RNase A variants engineered at the most destabilizing positions of the main hydrophobic core: further insight into protein stability. *Proteins Struct. Funct. Bioinforma.* 77(3):658–669.

58. Font J, Benito A, Lange R, Ribo M, Vilanova M (2006) The contribution of the residues from the main hydrophobic core of ribonuclease A to its pressure-folding transition state. *Protein Sci.* 15(5):1000–1009.

59. Frye KJ, Royer CA (1998) Probing the contribution of internal cavities to the volume change of protein unfolding under pressure. *Protein Sci.* 7(10):2217–22.

60. Spencer D, García-Moreno E. B, Stites WE (2013) The pH dependence of staphylococcal nuclease stability is incompatible with a three-state denaturation model. *Biophys. Chem.* 180-181:86–94.

61. Hamann S (1980) The role of electrostriction in high pressure chemistry.

62. Weber G (1992) *Protein interactions* (New York: Chapman and Hall).

63. Rasper J, Kauzmann W (1962) Volume Changes in Protein Reactions. 1. Ionization Reactions of Proteins. *J. Am. Chem. Soc.* 84(10):1771–1777.

64. Kauzmann W, Bodanszky A, Rasper J (1962) Volume Changes in Protein Reactions. 2. Comparison of Ionization Reactions in Proteins and Small Molecules. *J. Am. Chem. Soc.* 84(10):1777–1788.

65. Chalikian TV, Gindikin VS, Breslauer KJ (1998) Hydration of diglycyl tripeptides with non-polar side chains: a volumetric study. *Biophys. Chem.* 75(1):57–71.

66. Ehrhardt MR, Erijman L, Weber G, Wand AJ (1996) Molecular Recognition by Calmodulin : Pressure-Induced Reorganization of a Novel Calmodulin-peptide Complex. *Society* 35(5):1599–1605.

67. Brun L, Isom DG, Velu P, García-Moreno E. B, Royer CA (2006) Hydration of the Folding Transition State Ensemble of a Protein. *Biochemistry* 45(11):3473–3480.

68. Dwyer JJ, Gittis AG, Karp DA, Lattman EE, Spencer DS, Stites WE, García-Moreno E. B (2000) High apparent dielectric constants in the interior of a protein reflect water penetration. *Biophys. J.* 79(3):1610–1620.

69. Karp DA, Gittis AG, Stahley MR, Fitch CA, Stites WE, García-Moreno E. B (2007) High apparent dielectric constant inside a protein reflects structural reorganization coupled to the ionization of an internal Asp. *Biophys. J.* 92(6):2041–2053.

70. Privalov PL, Gill SJ (1987) Stability of protein structure and hydrophobic interaction. *Adv. Protein Chem.* 39:191–234.

71. Shortle D (1995) Staphylococcal nuclease: a showcase of m-value effects. *Adv. Protein Chem.* 46:217–247.

72. Durchschlag H (1986) in *Thermodynamic data for biochemistry and biotechnology* (Springer Berlin Heidelberg), pp 45–128.

73. Woolf L (1975) Tracer diffusion of tritiated water (THO) in ordinary water (H2O) under pressure. *J. Chem. Soc. Faraday Trans. 1 Phys. Chem. Condens. Phases* 71:784–796.

74. Herberhold H, Royer CA, Winter R (2004) Effects of chaotropic and kosmotropic cosolvents on the pressure-induced unfolding and denaturation of proteins: An FT-IR study on staphylococcal nuclease. *Biochemistry* 43(12):3336–3345.

75. Ravindra R, Royer CA, Winter R (2004) Pressure perturbation calorimetic studies of the solvation properties and the thermal unfolding of staphylococcal nuclease. *Phys. Chem. Chem. Phys.* 6(8):1952–1961.

76. Schweiker KL, Fitz VW, Makhatadze GI (2009) Universal convergence of the specific volume changes of globular proteins upon unfolding. *Biochemistry* 48(46):10846–10851.

77. Dellarole M, Roumestand C, Royer CA, Lecomte JTJ (2013) Volumetric properties underlying ligand binding in a monomeric hemoglobin: a high-pressure NMR study. *Biochim. Biophys. Acta* 1834(9):1910–22.

78. Rouget J-B, Aksel T, Roche J, Saldana J-L, Garcia AE, Barrick D, Royer CA (2011) Size and sequence and the volume change of protein folding. *J. Am. Chem. Soc.* 133(15):6020–6027.

79. Cellitti J, Llinas M, Echols N, Shank E a, Gillespie B, Kwon E, Crowder SM, Dahlquist FW, Alber T, Marqusee S (2007) Exploring subdomain cooperativity in T4 lysozyme I: structural and energetic studies of a circular permutant and protein fragment. *Protein Sci.* 16(5):842–851.

80. Cellitti J, Bernstein R, Marqusee S (2007) Exploring subdomain cooperativity in T4 lysozyme II: uncovering the C-terminal subdomain as a hidden intermediate in the kinetic folding pathway. *Protein Sci.* 16(5):852–862.

81. Llinás M, Marqusee S (1998) Subdomain interactions as a determinant in the folding and stability of T4 lysozyme. *Protein Sci.* 7(1):96–104.

82. Kitamura S, Sturtevant JM (1989) A scanning calorimetric study of the thermal denaturation of the lysozyme of phage T4 and the Arg 96 -> His mutant form thereof. *Biochemistry* 28(9):3788–3792.

83. Sturtevant JM (1987) Biochemical applications of differential scanning calorimetry. *Annu. Rev. Phys. Chem.* 38(1):463–488.

84. Pace CN, Scholtz JM (1997) Measuring the conformational stability of a protein. *Protein Struct. A Pract. approach* 2:299–321.

85. Liu R, Baase WA, Matthews BW (2000) The introduction of strain and its effects on the structure and stability of T4 lysozyme. *J. Mol. Biol.* 295(1):127–145.

86. Karpusas M, Baase WA, Matsumura M, Matthews BW (1989) Hydrophobic packing in T4 lysozyme probed by cavity-filling mutants. *Proc. Natl. Acad. Sci. U.S.A.* 86(21):8237–8241.

87. Eriksson AE, Baase WA, Matthews BW (1992) Similar hydrophobic replacements of Leu99 and Phe153 within the core of T4 lysozyme have different structural and thermodynamic consequences. *J. Mol. Biol.* 229(3):747–769.

88. Tanaka M, Chon H, Angkawidjaja C, Koga Y, Takano K, Kanaya S (2010) Protein core adaptability: crystal structures of the cavity-filling variants of Escherichia coli RNase HI. *Protein Pept. Lett.* 17(9):1163–9.

89. Wynn R, Harkins PC, Richards FM, Fox RO (1996) Mobile unnatural amino acid side chains in the core of staphylococcal nuclease. *Protein Sci.* 5(6):1026–1031.

90. Wynn R, Anderson CL, Richards FM, Fox RO (1995) Interactions in nonnative and truncated forms of staphylococcal nuclease as indicated by mutational free energy changes. *Protein Sci.* 4(9):1815–1823.

91. Wynn R, Harkins PC, Richards FM, Fox RO (1997) Comparison of straight chain and cyclic unnatural amino acids embedded in the core of staphylococcal nuclease. *Protein Sci.* 6(8):1621–6.

92. Fraser JS, Clarkson MW, Degnan SC, Erion R, Kern D, Alber T (2009) Hidden alternative structures of proline isomerase essential for catalysis. *Nature* 462(7273):669–73.

93. Lang PT, Ng H-L, Fraser JS, Corn JE, Echols N, Sales M, Holton JM, Alber T (2010) Automated electron-density sampling reveals widespread conformational polymorphism in proteins. *Protein Sci.* 19(7):1420–31.

94. Fraser JS, van den Bedem H, Samelson AJ, Lang PT, Holton JM, Echols N, Alber T (2011) Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci. U.S.A.* 108(39):16247–16252.

95. Millet O, Muhandiram DR, Skrynnikov NR, Kay LE (2002) Deuterium spin probes of side-chain dynamics in proteins. 1. Measurement of five relaxation rates per deuteron in (13)C-labeled and fractionally (2)H-enriched proteins in solution. *J. Am. Chem. Soc.* 124(22):6439–6448.

96. Skrynnikov NR, Millet O, Kay LE (2002) Deuterium spin probes of side-chain dynamics in proteins. 2. Spectral density mapping and identification of nanosecond time-scale side-chain motions. *J. Am. Chem. Soc.* 124(22):6449–60.

97. Mittermaier A, Kay LE, Forman-Kay JD (1999) Analysis of deuterium relaxation-derived methyl axis order parameters and correlation with local structure. *J. Biomol. NMR* 13(2):181–185.

98. Ming D, Brüschweiler R (2004) Prediction of methyl-side chain dynamics in proteins. *J. Biomol. NMR* 29(3):363–368.

99. Lee AL, Kinnear SA, Wand AJ (2000) Redistribution and loss of side chain entropy upon formation of a calmodulin-peptide complex. *Nat. Struct. Biol.* 7(1):72–77.

100. Lee AL, Wand AJ (2001) Microscopic origins of entropy, heat capacity and the glass transition in proteins. *Nature* 411(6836):501–4.

101. Igumenova TI, Frederick KK, Wand AJ (2006) Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution. *Chem. Rev.* 106(5):1672–99.

102. Lee AL, Sharp KA, Kranz JK, Song X-J, Wand AJ (2002) Temperature dependence of the internal dynamics of a calmodulin-peptide complex. *Biochemistry* 41(46):13814–13825.

103. Fuentes EJ, Der CJ, Lee AL (2004) Ligand-dependent Dynamics and Intramolecular Signaling in a PDZ Domain. *J. Mol. Biol.* 335(4):1105–1115.

104. Fuentes EJ, Gilmore SA, Mauldin RV, Lee AL (2006) Evaluation of energetic and dynamic coupling networks in a PDZ domain protein. *J. Mol. Biol.* 364(3):337–51.

105. Maeno A, Sindhikara D, Hirata F, Otten R, Frederick W, Yokoyama S, Akasaka K, Mulder F (2014) Cavity as a source of conformational fluctuation and high-energy state : High-pressure NMR study of a cavity-enlarged mutant of T4 lysozyme. *Biophys. J.*

106. Kamatari YO, Yamada H, Akasaka K, Jones JA, Dobson CM, Smith LJ (2001) Response of native and denatured hen lysozyme to high pressure studied by 15 N / 1 H NMR spectroscopy. 1793:1782–1793.

107. Bouvignies G, Vallurupalli P, Hansen DF, Correia BE, Lange O, Bah A, Vernon RM, Dahlquist FW, Baker D, Kay LE (2011) Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* 477(7362):111–114.

108. Hummer G, Garde S, García AE, Paulaitis ME, Pratt LR (1998) The pressure dependence of hydrophobic interactions is consistent with the observed pressure denaturation of proteins. *Proc. Natl. Acad. Sci. U.S.A.* 95(4):1552–1555.

109. Quillin ML, Breyer W a, Griswold IJ, Matthews BW (2000) Size versus polarizability in protein-ligand interactions: binding of noble gases within engineered cavities in phage T4 lysozyme. *J. Mol. Biol.* 302(4):955–977.

110. Wolfenden R, Radzicka A (1994) On the probability of finding a water molecule in a nonpolar cavity. *Science* 265(5174):936–937.

111. Vaitheeswaran S, Yin H, Rasaiah JC, Hummer G (2004) Water clusters in nonpolar cavities. *Proc. Natl. Acad. Sci. U.S.A.* 101(49):17002–17005.

112. Nucci NV, Pometun MS, Wand AJ (2011) Site-resolved measurement of water-protein interactions by solution NMR. *Nat. Struct. Mol. Biol.* 18(2):245–249.

113. Nucci NV, Pometun MS, Wand AJ (2011) Mapping the hydration dynamics of ubiquitin. *J. Am. Chem. Soc.* 133(32):12326–12329.

114. Damjanović A, García-Moreno E. B, Lattman EE, García AE (2005) Molecular dynamics study of hydration of the protein interior. *Comput. Phys. Commun.* 169(1-3):126–129.

115. Damjanović A, García-Moreno E. B, Lattman EE, García AE (2005) Molecular dynamics study of water penetration in staphylococcal nuclease. *Proteins Struct. Funct. Bioinforma.* 60(3):433–49.

116. García AE, Hummer G (2000) Water penetration and escape in proteins. *Proteins Struct. Funct. Bioinforma.* 38(3):261–72.

117. Damjanović A, Schlessman JL, Fitch CA, García AE, García-Moreno E. B (2007) Role of flexibility and polarity as determinants of the hydration of internal cavities and pockets in proteins. *Biophys. J.* 93(8):2791–804.

118. Sarupria S, Ghosh T, García AE, Garde S (2010) Studying pressure denaturation of a protein by molecular dynamics simulations. *Proteins Struct. Funct. Bioinforma.* 78(7):1641–51.

119. Persson F, Halle B (2013) Transient access to the protein interior: simulation versus NMR. *J. Am. Chem. Soc.* 135(23):8735–48.

120. Day R, Garcia AE (2008) Water penetration in the low and high pressure native states of ubiquitin. *Proteins Struct. Funct. Bioinforma.* 70(4):1175–1184.

121. Dill KA, Shortle D (1991) Denatured states of proteins. *Annu. Rev. Biochem.* 60:795–825.

122. Panick G, Vidugiris GJ a, Malessa R, Rapp G, Winter R, Royer C a (1999) Exploring the Temperature - Pressure Phase Diagram of Staphylococcal Nuclease †. 4157–4164.

123. Hilser VJ, García-Moreno E. B, Oas TG, Kapp G, Whitten ST (2006) A statistical thermodynamic model of the protein ensemble. *Chem. Rev.* 106(5):1545–1558.

124. Frye KJ, Royer CA (1998) Probing the contribution of internal cavities to the volume change of protein unfolding under pressure. *Protein Sci.* 7(10):2217–2222.

125. Roche J, Caro JA, Dellarole M, Guca E, Royer CA, García-Moreno E. B, Garcia AE, Roumestand C (2012) Structural, energetic, and dynamic responses of the native state ensemble of staphylococcal nuclease to cavity-creating mutations. *Proteins Struct. Funct. Bioinforma.* 135(39):14610–14618.

126. Roche J, Dellarole M, Caro JA, Guca E, Norberto DR, Yang Y, Garcia AE, Roumestand C, García-Moreno E. B, Royer CA (2012) Remodeling of the Folding Free Energy Landscape of Staphylococcal Nuclease by Cavity-Creating Mutations. *Biochemistry* 51(47):9535–9546.

127. Roche J, Dellarole M, Caro JA, Norberto DR, Garcia AE, García-Moreno E. B, Roumestand C, Royer CA (2013) Effect of Internal Cavities on Folding Rates and Routes Revealed by Real-time Pressure-Jump NMR Spectroscopy. *J. Am. Chem. Soc.* 82(6):1069–1080.

128. Sanner MF, Olson AJ, Lumire STF (1996) Reduced Surface: An Efficient Way to Compute Molecular Surfaces. *Biopolymers* 38:305–320.

CHAPTER 2:

EFFECT OF CAVITIES ON THE THERMODYNAMIC STABILITY AND PRESSURE UNFOLDING OF PROTEINS

Abstract

According to LeChâtelier's principle, high pressure unfolds proteins because the volume of the unfolded state is smaller than the volume of the folded state. Many physical factors have been invoked to explain the difference in volume between folded and unfolded state (ΔV). Recent studies demonstrate that ΔV reflects primarily the volume of internal cavities that are present in the native state and absent in the unfolded one. This was the conclusion from extensive pressure-unfolding measurements with ten variants of staphylococcal nuclease (SNase) with artificial cavities engineered through substitution of core positions with Ala. The data from that study did not cover a large enough range of ΔV to allow establishing fundamental relationships between ΔV and the cavity volume observed in crystal structures or with thermodynamic stability (ΔG^o). Here, these problems are surmounted by including a new set of variants of SNase in which 2 or 3 internal positions were substituted with Ala. Crystal structures of all variants were obtained. Pressure unfolding was monitored with Trp fluorescence or NMR spectroscopy. These proteins reported large increases in ΔV, consistent with what was observed in the crystal structures. With data from a total of 20 cavity-containing variants of SNase it was possible to establish a linear relationship between the cavity volumes calculated from the structures and the measured ΔV. The data are all consistent with a role of the cavities in folded state as the dominant determinant of pressure unfolding of SNase and possibly of all proteins. Site-specific ΔV values obtained with NMR spectroscopy showed that the backbone amides that report the largest ΔV values cluster near the cavities. A linear relationship between ΔG^o measured independently with chemical denaturation and ΔV measured by pressure unfolding shows that generating a cavity in the protein costs 80

cal/mL. The fundamental relationships established here should enable structure-based prediction of volumes and pressure sensitivity of proteins.

Introduction

Proteins are densely packed, with densities comparable to that of a crystalline solid (1–3). Nevertheless, cavities large enough to hold water molecules can be found in the interior of proteins (4). In the crystal structures, these cavities are usually dry, totally devoid of water molecules (5–7). They are thought to be functionally important to regulate dynamics in the protein interior needed for function (8, 9) and for the evolvability of function (10, 11). It was recently proposed that the volume of these cavities is also largely responsible for the difference in volume between folded and unfolded states, ΔV**,** and thus for the unfolding of proteins by high hydrostatic pressures (12, 13). However, the exact relationship between cavity volumes observed in crystal structures and ΔV measure experimentally has yet to be established. Until then, contributions to ΔV from factors like electrostriction (14), compressibility (15) and expansivity (16, 17) cannot be ruled out.

Previously we engineered ten cavity-containing variants of a highly stable form of staphylococcal nuclease (SNase) known as Δ+PHS by substitution of internal positions with Ala (V23A, L25A, F34A, L36A, L38A, V66A, V74A, I92A, L103A and L125A) (12). The goal was to evaluate systematically the role of cavities as determinants of the pressure sensitivity of proteins. Crystal structures of the variants revealed increases in the volume of cavities due to the substitutions. The cavities are all observed to be dry, with no evidence of internal water molecules. They have no detectable impact on any other structural aspect of the protein. Pressure unfolding experiments monitored with Trp fluorescence and with ^{1}H-^{15}N HSQC spectra reported increases in ΔV for all of the variants studied, relative to the background protein. A single substitution to alanine in the

I92A variant doubled the original ΔV of Δ+PHS SNase from 58 mL/mol to 113 mL/mol. Unfortunately, even with this extensive set of 10 proteins with different cavity volumes, it was not possible to determine a quantitative relationship between the volume of cavities observed in crystal structures and the ΔV measured from pressure unfolding. The problem could stem from biases in the crystal structure, but the real problem is that in a plot of measured vs. calculated ΔV the values are not distributed over a wide enough range of ΔV to make fitting with a linear model meaningful. Understanding the relationship between cavities observed in crystal structures and ΔV measured experimentally is necessary to explain quantitatively the role of cavities as a determinants of the pressure unfolding of proteins, to allow structure-based prediction of pressure sensitivity of proteins, and to determine if other factors contribute towards the ΔV responsible for pressure unfolding.

Several physical factors other than cavity volume have been proposed to explain the difference in volume between folded and unfolded proteins. The exposure of buried surface area to solvent upon unfolding, which determines the heat and chemical denaturation of proteins, results in changes in solvation volumes that are thought to contribute significantly to the pressure sensitivity of proteins (14, 18–20). Changes in the charge state of ionizable residues upon unfolding could have a very large effect on solvation volumes. The compressibility (15, 21, 22) and expansivity (17, 23–25) of proteins are second order factors that could change in response to the engineered substitutions. Although compression of proteins is readily observed, the isothermal compression of proteins is small, about ten times smaller than that of ice, and no clear change in compressibility between folded and unfolded states has been measured to date.

Temperature dependent changes in ΔV are reflected in the expansivity and its interpretation is challenged by the convolution of pressure and temperature effects on volume changes (26). Co-solvents like guanidine (GdnHCl) are also frequently used to modulate protein stability and to lower the mid-point of the pressure unfolding reaction to the experimentally accessible pressure range. No detectable effect of GdnHCl on ΔV has been measured to date, although other chaotropes, including urea and methanol, can have measurable effects on the pressure behavior of proteins (27–31). Computational studies emphasize the importance of pressure-dependent changes in solvent structure that could lead to the destabilization of proteins under pressure, but this cannot be validated experimentally (32). In summary, several physical properties of proteins and of water could be responsible for the pressure sensitivity of proteins, but the only contribution that has been validated experimentally beyond reasonable doubt is the large contribution related to internal cavities in the folded state.

To determine the relationship between the volume of cavities observed in crystal structures and the ΔV of unfolding measured experimentally, the range of cavity volumes of SNase variants was expanded by use of as many as three simultaneous Ala substitutions in the hydrophobic core. Because the I92A variant reports the highest ΔV out of the ten variants previously studied, it was used as the starting point to engineer variants with multiple alanine substitutions, combining the I92A substitution with substitutions V23A, L25A, L36A, V66A, V74A, L103A, L125A, L25A/L36A and L25A/V66A. Although cavity formation is always destabilizing to proteins, the proteins tolerated the Ala substitutions without unfolding. This is possible because of the high stability of the parent protein, Δ+PHS, which was engineered to have 11.9 kcal/mol of

stability. Crystal structures of the nine new variants were determined and used in volume calculations. ΔV was measured by pressure unfolding monitoring with Trp fluorescence. Thermodynamic stability was determined independently by GdnHCl denaturation monitored with Trp fluorescence. The structural and thermodynamic data collected with the new variants, together with the data previously collected for Δ+PHS and its ten variants with a single alanine substitution, allowed an unprecedented and systematic evaluation of the relationship between ΔV measured from pressure unfolding experiments and the volume of cavities observed in crystal structures. The data also described the relationship between ΔG^{o}_{H2O} and ΔV, which reports on the thermodynamic cost associated with the creation of cavities in the interior of a protein. These two fundamental thermodynamic properties of proteins should prove useful for structure-based predictions of the energetic penalties of cavity engineering.

The local response of proteins to the application of high hydrostatic pressure was also examined using high pressure NMR spectroscopy. The pressure unfolding of the V66A/I92A, I92A/L125A and L25A/V66A/I92A variants was monitored by following cross-peak intensities of the more than 80% of amides recorded by ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectra. The variants with the largest cavities exhibited clearly detectable changes in the local response to pressure that were previously difficult to discern with variants with a single Ala substitution. The volumetric data with atomic-resolution for variants with very large cavities allowed identification of the regions of the protein with the largest ΔV and the largest ΔΔV, ($\Delta V_{variant}$ - $\Delta V_{background}$). These were mapped onto the crystal structures to determine the relationship to the location of the cavities.

Results

Crystal structures. All crystal structures were obtained at 1 bar and under cryogenic conditions (Fig. 1 and Suppl. Table 1). Superposition of the structures onto the parent protein using LSQKAB (33) revealed no significant differences in the backbone traces (RMSD with respect to parent protein < 0.2 Å). Only few changes were observed in side chain rotameric states near the region adjacent to the engineered cavity, usually changes that would minimize the added volume to the cavity. The structures revealed the enlargement of the natural microcavity or in the case of I92A/L103A and I92A/L125A the presence of multiple, large internal cavities. The largest cavity was seen in the L25A/L36A/I92A variant, in which nearly the entire β-barrel was hollowed out. No crystallographic water molecules were observed within any of the engineered cavities. No electron density was found within the cavities even at the lowest contour level that was examined.

Calculated cavity volume. Many algorithms for volume calculations in proteins use a probe the size of a water molecule to define the protein surface (34–36). The contact point between the probe and the van der Waals surface of the protein is defined as the contact surface (3). The distance from the center of the probe to the protein surface is known as the probe accessible surface, since a probe can be placed without any steric clashes in any volume identified by the probe (3). The contact point between the probe and the protein surface defines the so-called molecular surface (3). The main purpose of using such a probe is to define a smooth layer that ignores the spaces between atoms, or

interstitial volumes, that are too small to fit a probe. The surface seals off the interior of the protein from the outside and any internal volumes large enough to fit a probe are identified as cavities. Smaller probes will identify larger and more numerous cavities. When the probe is too small the surface connects the interior surface to the outer surface, turning the cavities into crevasses accessible from the outside.

The use of probes to define surfaces leads to a surface that obscures the complex topology of proteins and to overly-sensitive algorithms in which small changes in atomic coordinates can result in the difference between the opening and closing of a crevasse or the appearance and disappearance of cavities. Because of the difficult problems inherent to the definition of internal cavities and to the calculation of their volumes, we compared the performance of three different algorithms, CASTp (36), McVol (35) and VOIDOO (34), based on Delaunay triangulation, Montecarlo methods, and a numerical grid, respectively. The cavities identified by the algorithms were examined visually to confirm that they are near the substitution sites. Only cavities adjacent to the substitution site or clearly created in response to the substitution (Fig. 1) were included in the calculated cavity volumes in Suppl. Table 1. These are, indeed, the only cavities in the protein that are fully internal and not in contact with bulk water.

For the set of 20 proteins studied here, the cavity volumes measured by CASTp and a probe of 1.4 Å ranged from 30 mL/mol for variant L25A to 165 mL/mol for variant L25A/L36A/I92A (Suppl. Table 1). CASTp fails to recognize Δ+PHS as the protein with the smallest cavity. This is because the error of such calculations increases as the volume of the cavity approaches that of the probe. A probe of 1.4 Å will not consider a void

sphere of radius 1.39 Å, even though it is only 0.14 mL/mol below the smallest detectable cavity.

McVol was also used to calculate cavity volumes. This algorithm results in absolute volumes that are smaller than those obtained using CASTp. This is because the algorithms use inherently different principles for volume calculations and their output cannot be compared directly. To allow for an improved qualitative comparison between the two algorithms, volumes were also calculated using a probe of 1.2 Å. McVol, with a probe of 1.2 Å, identifies Δ+PHS as having the smallest cavity volume, namely 40 mL/mol. Variant L25A/L36A/I92A once again has the largest cavity of 150 mL/mol.

In comparison, VOIDOO performs poorly and does not detect cavities for most variants using a probe of 1.4 Å and requires a probe of 1.0 Å to find the cavities found by a probe of 1.4 Å and using CASTp or McVol. Furthermore, the volumes obtained using VOIDOO carry a very large error comparable to the value itself. Smaller probe radii were not used because they fail to separate the interior of the protein from bulk and under these conditions no cavities isolated from bulk can be detected.

One of the problems with volume calculations based on static crystal structures is that they are agnostic to the volumes associated with the fluctuations that might be experienced in the relatively long time scales of the equilibrium thermodynamic experiments used to measure ΔV. Molecular dynamics simulations of variants L36A/I92A and L25A/L36A/I92A were used to assess how the faster fluctuations of the protein would affect the measured cavity volume. The overall structure of the protein persists over 10 ns runs. Rotameric changes were observed for side chains lining the cavities, resulting in considerable fluctuations of the cavity volume, with standard

deviations of over 30% of the average volume of the cavities over 10 ns. The volume calculation only included cavities adjacent to the substitution sites, although transient cavities formed throughout the simulation. Transient cavities away from the substitution sites were observed in simulations of all variants. It is possible that these small fluctuations in volume give rise to the deviations of certain variants from a linear correlation with respect to the experimental ΔV (Fig. 6A and B). However, the fluctuations of the engineered cavities can account for most of the deviation from a linear correlation (Suppl. fig 4). Furthermore, the linear relationship between experimental ΔV and the van der Waals volume of the removed side chain suggests that the effect of the substitutions on the volume of transient cavities remains constant between variants (Fig. 6C). This implies that whatever the contribution from transient cavities is to the ΔV of the protein, it is present in all variants, including the parent protein, and should not affect the increase in ΔV due to the substitutions.

No water penetration into the cavity was observed throughout the simulations. This, however, does not come as a surprise. Damjanović et al. (37–39) showed that water penetration into the hydrophobic core of SNase depends heavily on the initial hydration state of the protein. Simulations that started with the protein already internally hydrated, either artificially or because the crystal structure already showed internal water molecules, resulted in many more events of water penetration and escape. Moreover, multiple simulations of the same protein proved much more efficient at sampling different hydration states than one long simulation. This is directly related to problems with convergence over the timescale of the simulation. The simulations by Damjanović et al. were performed on SNase variants that contained a buried polar or ionizable residue in

the protein core, which served as an attractor to water molecules. Without a polar side chain lining the cavity, the authors argued that water penetration into the protein core would be unlikely. The 10 ns simulations presented here started with a dry cavity (as observed in the crystal structure) lined by hydrophobic residues. The conformational fluctuations that took place throughout the run were simply not sufficient to accurately sample the hydration state of the engineered cavities.

Thermodynamic stability. The stability of Δ+PHS SNase and its variants was measured at pH 7, 298 K, 1 atm with guanidine hydrochloride (GdnHCl) denaturation monitored by Trp fluorescence and analyzed using a two-state model (40). As expected from previous work (12, 41) the stability of the Ala-substituted variants was always lower than the 11.9 kcal/mol of the parent protein (Fig. 2A and Table 2). The lowest stability was recorded for the L25A/L36A/I92A variant (ΔG^{o}_{H2O} = 0.4 kcal/mol). The energetic penalties due to an Ala substitution were within the expected range based on previous studies (4). No significant deviations from two-state behavior were observed, despite the fact that some variants suffered up to three substitutions to Ala in the same general region of the protein (Fig. 2A). This was not the case for less severe insults in T4 lysozyme variants (4). It is well known that T4 lysozyme deviates from two-state unfolding (42–45), while the behavior of SNase is inconsistent with three-state unfolding by pH (46).

It was not always possible to predict the stabilities of variants with multiple Ala substitutions by adding the energetic cost of engineering each of the variants with a single Ala substitution (Table 1 and Suppl. fig 2A). For example, variant L25A/I92A was within error of the value obtained by adding the energetic penalty of engineering the single

variants L25A and I92A. Variant L36A/I92A, on the other hand, was 2.2 kcal/mol less stable than expected from adding the costs of L36A and I92A. It is difficult to rationalize from the structural and thermodynamic information why this is the case. Both double Ala variants involve the same type of substitution, involve positions in the main core of the protein, and result in similar changes in ΔV (Table 1). Yet variant L36A/I92A was by far the least stable of the double Ala variants and its crystal structure displayed one of the largest cavities in this study (Suppl. table 2). Its marked deviation from additivity highlights the important energetic consequences that subtle differences in the location of the cavity can result in.

For the variants with two or three alanine substitutions, the *m*-value was always larger than the parent protein (*m* = 4.9 kcal/mol*M). The largest *m*-value was measured for the L25A/L36A/I92A variant (*m* = 6.4 kcal/mol*M), an increase in *m*-value by a factor of 1.31. This is consistent with the idea that *m*-values reflect the amount of backbone surface area exposed upon unfolding (47, 48). The truncations of core hydrophobic residues to Ala will not change the amount of polar backbone that is exposed to GdnHCl in the folded state, since the substitution sites are secluded from bulk solvent. In the unfolded state, however, there will be a decreased occlusion of the polar backbone surface by the larger hydrophobic side chain. This will result in an increase in polar backbone surface exposed to the denaturant, which will result in an increase in the cooperativity of the unfolding reaction (47).

Pressure unfolding monitored by Trp fluorescence. Pressure unfolding of all variants was monitored using the fluorescence of Trp-140 and analyzed with a two-state model

with a linear dependence of ΔG on pressure (Fig. 2B). All variants with artificial cavities engineered with substitutions with Ala report an increase in ΔV compared to the parent protein Δ+PHS SNase, which has a ΔV = 58 mL/mol (Fig. 2 and 5). Variants with two or three Ala substitution always include substitution Ala-92 and show an increase in ΔV compared to I92A (ΔV = 113 mL/mol) as well. The largest ΔV was measured for the L25A/L36A/I92A variant (ΔV = 170 mL/mol), 3 times the value of the parent protein. For a subset of variants, adding the increases in ΔV values of variants with a single Ala substitution adequately predicted the ΔV of variants with multiple substitutions (Suppl. fig 2B). These were variants that involved substitutions that generated cavities far apart in space, like I92A/L125A, or variants that involved Val-to-Ala substitutions, the type of substitution that results in the smallest change in volume. In contrast, variants such as L36A/I92A reported ΔV values that were significantly lower than the values expected by adding the changes observed for the individual substitutions. These variants contained the large side chain substitution Leu-to-Ala at positions that are proximal to position 92, indicating that there exists some overlap between the cavities generated by substitutions L25A and I92A. This shared interstitial volume will be included in the change in ΔV caused by the first substitution but not the second.

The instrumentation used to measure pressure unfolding tolerates pressure up to 3000 bar. For proteins that are too stable, including many variants of SNase, the pressure range is not sufficient to sample the entire unfolding curve; it would require more than 10 kbar to unfold Δ+PHS SNase. To circumvent this experimental limitation, different amounts of GdnHCl are used to lower the stability and to bring the transition into the range of pressures that is accessible. No detectable effects of GdnHCl on ΔV have been

measured to date, although other chaotropes, including urea and methanol, can have measurable effects on the pressure behavior of proteins (25, 28, 30, 31). Previously, variants of Δ+PHS with a single Ala substitution were unfolded at three different concentrations of GdnHCl (Table 2). The ΔG values recovered from an extrapolation to zero GdnHCl were always consistent with those obtained from GdnHCl unfolding experiments at 1 atm. However, the ΔV values of a given variant could differ by up to 10 mL/mol depending on the amount of GdnHCl present in solution. The fact that different concentrations of GdnHCl resulted in similar variations in ΔV argues against a direct effect of denaturant on the ΔV of the protein. In the present study, the variants V23A/I92A, L25A/I92A and V66A/I92A were pressure unfolded in as low as 0.35 M, 0.35 M and 0.4 M GdnHCl, respectively, and compared to their behavior in the absence of GdnHCl. Once again, the curves yielded ΔV values that differ by up to 20 mL/mol compared to their unfolding curves at 0 M GdnHCl.

Close inspection of the curves revealed a small but persistent deviation of the data from the fit near the denatured baseline just before the plateau for almost all variants (Fig. 2B). This deviation from two-state behavior could be due to residual structure in the denatured state that continues to partially occlude the Trp from bulk water. It could also arise from population of an intermediate state or from the need to include a term in the fitting model that accounts explicitly for changes in compressibility upon unfolding. Higher order models indeed reproduce the data, but yield very large uncertainties in the fitted values, making it impossible to distinguish which model is correct. Repeating the experiment under identical conditions showed that the deviation is reproducible, and that fitting with a two-state model could reliably yield the same ΔV. Such an artifact in the

denatured baseline of the fluorescence curves of a protein can result in the observed variability of ΔV values at varying GdnHCl concentration. Adding GdnHCl will shift the transition midpoint to lower pressures, allowing data collection to extend further into the denatured baseline, and possibly accounting for the variability in the ΔV value of a given protein. The data suggest that the effects of GdnHCl on ΔV, if any, are small compared to the resolution of the technique and to the measured changes in ΔV.

Pressure unfolding monitored by NMR spectroscopy. The pressure unfolding of Δ+PHS SNase and of the V66A/I92A, I92A/L125A and L25A/V66A/I92A variants was monitored by ^{1}H-^{15}N HSQC spectra by following the maximum intensity of crosspeaks until broadened beyond detection. With this approach the unfolding profiles of about 130 detectable amide groups were measured for each protein (Fig. 3 and Suppl. Fig. 1). The three variants that were studied with NMR spectroscopy were chosen based on their large ΔV as measured by fluorescence and based on the location of the cavities. Variants V66A/I92A and L25A/V66A/I92A introduce cavities in the β-barrel of SNase, while variant I92A/L125A generates cavities in both the β-barrel and where helix-2 and helix-3 meet. These substitutions probe both helical (position 66 and 125) and β-strand (position 25 and 92) elements of secondary structure and build on the NMR data collected previously with the I92A, V66A and L125A variants (12). In agreement with what is observed by crystallography, the overlay of HSQC spectra of the three variants studied here show only small, localized changes in chemical shifts near the substitution sites, suggesting that the overall structure of the protein was not affected by the Ala

substitution (Suppl. Fig. 2). The increase in hydrostatic pressure also does not lead to clear localized changes in chemical shifts.

Nearly 70% of the peaks displayed a large decrease in intensity. The intensity measured as a function of pressure was fitted using a two-state model. The average ΔV value (and standard deviation) for the V66A/I92A, I92A/L125A and L25A/V66A/I92A variants was 144 (± 25), 192 (± 22), and 201 (± 27) mL/mol, respectively. These values are systematically larger than those from fluorescence, although the value reported by Trp-fluorescence falls within the distribution of values reported by NMR spectroscopy, as previously observed. The ΔV measured by Trp fluorescence for the V66A/I92A and L25A/V66A/I92A variants compares well with the value obtained from the resonance of the Trp-140 indole. In contrast the ΔV measured from the indole resonance in the I92A/L125A variant is significantly larger than the value obtained by fluorescence (Table 2). The L125A substitution generates a cavity between helix-2 and helix-3 of SNase that could increase the hydration of Trp-140, located in the capping motif at the end of helix-3 (49). This could quench the fluorescence intensity in the folded state and thus decrease the ΔV reported by fluorescence of I92A/L125A.

The ΔV values measured with NMR spectroscopy were on average larger than the ΔV of Δ+PHS of 89 (± 13) or for I92A of 130 (± 21) (Fig. 5 and Table 3), consistent with the trend observed with Trp fluorescence. In the case of the L25A/V66A/I92A variant the average ΔV increased by a factor of two compared to the Δ+PHS parent protein. The differences ($\Delta\Delta V = \Delta V_{variant} - \Delta V_{parent}$), ranged from 45 mL/mol for position 26 to 169 mL/mol for position 90, both nearby substitution sites. No site reported a negative ΔΔV except sites 26 and 33 in variant V66A/I92A. For each variant, the ΔV values are

distributed around a mean value that is the most frequent value reported by the protein (Fig. 3E-F). The width of the distribution changes depending on the variant, and for variant L25A/V66A/I92A appears almost bimodal (Fig. 3F). This is consistent with the fact that for this variant almost all large ΔV values map onto the β-barrel region, while helix-3 reports average or lower than average ΔV values. It is possible that the structural bias in pressure sensitivity, caused by the insult of three substitutions to Ala in the same region, begins to reflect the independent unfolding behavior of the subdomains in SNase. Two outliers with much larger ΔV values than the rest can be seen in the histogram for variant V66A/I92A, and correspond to residues Ala-12 and His-121 (Fig. 3F). The intensity profiles and chemical shifts of these variants do not show any features that would set them apart. They are located at the beginning of β-strand-1 and in the capping motif of helix-3, respectively. It is unlikely that these residues act as major structural lynch pins in the pressure unfolding of SNase because they do not stand out in the other two variants (I92A/L125A and L25A/V66A/I92A). It is difficult to extract the significance of these outliers from the present data.

A subset of the unfolding profiles measured from pressure sensitivity of amides (19 out of 89 amides for the I92A variant in 0.65 M GdnHCl, 70 out of 98 amides for the V66A/I92A variant, 48 out of 83 amides for the L25A/V66A/I92A variant) show a clear increase in maximum intensity of the crosspeak in the pre-transition region that plateaus at a specific pressure (800 bar for I92A in 0.65M GdnHCl, 800 bar for V66A/I92A and 350 bar for L25A/V66A/I92A). Two-state fitting to obtain an estimate of the thermodynamics of the main transition was performed starting from the pressure at which the majority of curves attain a maximum value, i.e., the plateau value stated above. This

reduced the number of data points in the fit, which increased fitting error. To evaluate the effect of reducing the number of data points, curves that do not show the increase in intensity were fitted with and without the initial data points and the effect on the fitted values remained within the error. It is worth noting that a compressibility model and a three-state model successfully fit the data with pre-transition increases in intensity, but this ambiguity and the non-physical values obtained from the fits barred any further application for our intended purposes.

Discussion

The variants of Δ+PHS with one, two or three internal groups substituted with Ala were destabilized relative to the parent protein, by as much as 11 kcal/mol in the case of the L25A/L36A/I92A variant (Fig. 2 and Table 1). The crystal structures show that the substitutions hollow out the β-barrel in the OB-fold or generate more than one large cavity without any detectable effects on the global structure of the protein (Fig. 1). Pressure unfolding experiments of the variants show how the ΔV increases with the number of alanine substitutions. The L25A/L36A/I92A variant shows the largest cavity volume, regardless of the algorithm used to calculate the volume. With CASTp using a probe of 1.4 Å its cavity has a volume of 165 mL/mol (Suppl. Table 2). This represents a three-fold increase in ΔV compared to the parent protein (Fig. 2 and Table 1).

Physical origins of pressure unfolding. With the set of variants with single Ala substitutions it was not possible to determine the relationship between the measured ΔV and the cavity volume calculated from structures (12). The problem lies in that the scatter

inherent to the data is comparable to the range of ΔV covered by the data. This problem has now been superseded by the addition of data from proteins with two or three Ala substitutions. The measured ΔV plotted against the volumes calculated with the CASTp algorithm using a probe of 1.4 Å (or 1.2 Å for McVol) display a linear relationship (Fig. 6).

Because volume calculations are highly sensitive to the probe size used, the error of such calculations increases as the volume of the cavity approaches that of the probe. Similarly, increasing the probe size leads to a surface that is less sensitive to the topology of proteins and cavities. The larger the probe, the larger the interstitial volume that is ignored, defined as the unavoidable empty space that exists in a lattice of packed spheres and that cannot fit another sphere. Thus, as cavities approach the volume of the probe, the calculation will increasingly underestimate volume, leading to a deviation from a straight line (Fig. 6A). Reducing the probe size to 1.2 Å reduces the deviation considerably but also overestimates cavity volumes, as evidenced by the displacement of all points toward the right of the diagonal (Fig. 6B).

Probe-based algorithms are inherently biased by the choice of probe size. A less biased measure of the cavity volume generated by a substitution can be obtained from the difference in van der Waals volume of the original side chain and Ala. This method is insensitive to structural responses of the protein to the substitution and thus might overestimate the volume of the generated cavity. The low RMSD between the structures of variants and the parent protein, as well as the little to no change in rotameric state of residues adjacent to the cavity, supports the use of this approach. The largest outlier to the right of the diagonal corresponds to the F34A variant, the variant with the largest

backbone RMSD as compared to the parent protein. In contrast, most points deviate slightly to the left of the diagonal, with the largest outlier corresponding to the I92A variant. This underestimation of the change in cavity volume probably reflects the additional volume in the cavities generated by the truncations to Ala, which include the volume from the truncated side chain as well as the interstitial volume between the side chain and the rest of the protein. Because the parent protein is used as reference, it serves as the origin of the abscissa. Plotting the ΔV measured by fluorescence against the change in the van der Waals volume of the side chains yields points that fall remarkably close to a line of unity slope (Fig. 6C). Overall the linear relationship between the experimentally measured ΔV and the cavity volume calculated from structure suggests that no other factors other than cavities need to be invoked to interpret the measured ΔV. The data suggest that cavities can account for the entire ΔV of SNase. Contributions from electrostriction, changes in expansivity, and compressibility, must be negligible.

A linear relationship between ΔG^o and ΔV. The consequences of cavity-creating substitutions on thermodynamic stability ($\Delta\Delta G^o = \Delta G^o_{variant} - \Delta G^o_{background}$) were measured for the 19 variants with one, two or three alanine substitutions and compared to the measured ΔV. The points suggest a linear relationship with a slope of 0.08 ± 0.008 kcal/mL (Fig. 7). The absence of any clear curvature and the small residuals support the simplified model used to analyze the pressure unfolding data that does not include an explicit term for compressibility (21). The linear correlation shows an intercept at -3.7 ± 0.9 kcal/mol, which represents the energetic penalty related to the net cavity volume of the Δ+PHS background and corresponds to a ΔV of 44 ± 10 mL/mol. This is close to the

58 mL/mol measured by fluorescence, which is remarkable considering that a total of 20 proteins compose the fit. The cavity observed in the structure of Δ+PHS (Fig. 1A) must be responsible for at least part of the 44 mL/mol. This indicates that it should be possible to reduce the ΔV of Δ+PHS by engineering small-to-large substitutions that fill the cavity. However, this approach can probably not eliminate the pressure sensitivity of Δ+PHS, since other small packing defects throughout the protein must also contribute to the ΔV.

The slope of 0.08 kcal/mL is about twice the value reported by Eriksson et al. (4, 41) of 0.04-0.05 kcal/mL, obtained by correlating thermodynamic stability of T4 lysozyme variants to cavity volumes calculated from crystal structures. There are two main reasons for this discrepancy. The correlation drawn by Eriksson et al. related the ΔGº of T4 lysozyme variants to the volume of cavities observed in the crystal structures. This type of analysis is strongly dependent on the algorithm and probe size used (Fig. 6 and Suppl. fig 4) and the treatment of crystallographic waters observed (13). Here, we relate two thermodynamic values measured experimentally (Fig. 7). The work by Eriksson et al. was done on T4 lysozyme, a protein with two distinct domains and that is known to deviate from two-state unfolding (42–45, 50). This undermines the thermodynamic values obtained by the authors. In contrast, SNase is known to unfold following a two-state model in a wide range of conditions (51). It is possible, nonetheless, that engineering cavities as large as the ones studied here could result in a non-cooperative unfolding behavior. The histograms of site-specific ΔV values obtained by NMR spectroscopy indicate that engineering cavities can also affect the width of the

distribution (Fig. 3). However, the distributions remain within the envelope of a single Gaussian distribution.

The slope of 0.08 kcal/mL implies that the formation of a cavity of 12.5 mL/mol, large enough to accommodate a water molecule, would destabilize a protein by 1 kcal/mol. The thermodynamic relationship established between these two experimental values, between thermodynamic stability measured by chemical denaturation, ΔG^{o}, and the apparent cavity volume measured by pressure unfolding, ΔV, will be useful for structure-based predictions of the relationship of changes in cavity volumes and stability.

The effect of cavities on ΔV is location dependent. The residuals from a linear regression analysis of the plot of ΔV versus cavity volumes suggest that the same type of substitution can result in different cavity volumes depending on where in the structure the cavity is generated. Comparison of variants with two and three alanine substitution against variants with a single alanine substitution shows an additive trend for five of the nine variants studied here (Suppl. Fig. 6). Interestingly, they involve substitutions that are either far apart in space and generate cavities that do not connect, or they involve truncations of Val, the smallest change possible. The four outliers always report a lower ΔV than expected and involve cavities that overlap and truncations of the large side chains of Leu or Ile. Cavities that overlap are expected to contribute less to ΔV when they are generated simultaneously because the change in ΔV due to the first substitution will already have accounted for the overlap volume by the time the second substitution is introduced. This highlights the fact that interstitial volumes between closely packed

protein atoms account for part of the increase in ΔV due to the cavity generated by the truncation to alanine, and also to the baseline ΔV measured in the parent protein.

Location also determines the response of the structure to the generation of a cavity. In a more rigid part of the protein the cavity generated by a substitution to Ala will be larger than in a more flexible part of the protein, where the flexibility will facilitate compensation for the empty volume left behind by the substitution (52). For example, the structures of variants F34A and V66A/I92A show backbone atoms with high RMS deviation from Δ+PHS localized near the substitution sites, namely in the β-2 to β-3 turn and after helix-2, respectively. These structures also show rotameric changes at positions 23, 32 and 36 for F34A and 36 and 74 for V66A/I92A that move the side chain toward the cavity. NMR spectroscopy also shows structural responses that are localized to the region near the substitution sites, observed as changes in chemical shifts between the variants and Δ+PHS (Suppl. Fig. 3).

Different regions of the protein experience fluctuations of different magnitude. This is not reflected in crystal structures. These fluctuations determine the range of volumes of the cavities in solution, which might be very different from the volumes of the cavities observed in a static crystal structure. To evaluate the effect of local fluctuations on the volume of cavities, 10 ns simulations were run on the L36A/I92A and L25A/L36A/I92A variants. The volumes of the cavities were monitored throughout the trajectory. A standard deviation of about 25% of the mean value was observed. This provides a sense of how large the fluctuations in the volume of a cavity in a protein in solution can be. The mean cavity volume averaged over the 10 ns trajectory, on the other hand, is close to the value obtained from the crystal structure, suggesting that the cavity

volumes measured from crystal structures are useful for the relationships established by the data in Fig. 6 and Suppl. Fig. 4.

The local structural consequences of the application of high hydrostatic pressure were further inspected by NMR spectroscopy for the V66A/I92A, I92A/L125A, L25A/V66A/I92A variants. The pressure unfolding process was monitored by following the peak intensities of amide resonances in ^{15}N-^{1}H HSQC NMR spectra (Fig. 3 and Suppl. Fig. 1). The mean value of nearly 90 site-specific ΔV values increased with increasing cavity volume, consistent with the findings from Trp fluorescence experiments. The structural mapping of these values shows a heterogeneous distribution of pressure sensitivity throughout the protein (Fig. 3). Most notably, the largest increases in site-specific ΔV in Ala-substituted variants compared to the Δ+PHS parent protein are clustered around the cavity generated by the substitutions. This is consistent with the idea that the pressure sensitivity of proteins is determined by the distribution of cavities. The site-specific ΔV show a rather wide range of values (Fig. 4) that might be direct measures of the local volumetric fluctuations of the protein.

Sites adjacent to the engineered cavities report an overall increase in ΔV in all three variants. The cavity at position 125 in variant I92A/L125A induces large increases in the ΔV reported by amides on helix-3 (Fig. 4B), which in the other two variants reports only moderate to low increases in ΔV. Helix-2 is in contact with both the β-barrel and helix-3 of SNase and experiences some large increases in ΔV regardless of where the cavity is introduced (Fig. 4D-F). The strand at the top of the β-barrel, β-strand-3, also reports large increases in ΔV with respect to the parent protein, probably because of its proximity to substitution I92A which is present in all variants. In the case of variant

L25A/V66A/I92A, portions of all but one β-strand and almost the entire helix-1 show the largest increases in ΔV, which propagates into the C-terminal of the protein (Fig. 4F). This strong response to pressure in the region where Trp-140 is located could reflect how pressure induced changes in local fluctuations and hydration affect the fluorescence of the Trp side chain and lead to the discrepancy with respect to the ΔV values obtained by fluorescence.

Recent studies show that, even in the small protein ubiquitin, hydrophobic side chains in the protein core display very different dynamics as a function of pressure (53). Some side chains show dynamics restrained to a single rotamer but neighboring side chains can be observed exploring multiple rotameric conformations (53). The clustering of such motions through the hydrophobic core of proteins has been proposed as an allosteric mechanism of proteins (9, 54, 55). Simulations of SNase and its variants with internal polar groups indicate that specific routes exist for water penetration into the protein core (37, 38, 56). The local fluctuations and packing defect that enable transient water access into the protein core could give rise to the most pressure sensitive regions of the protein highlighted by the site specific ΔV values.

Pre-denaturation response to pressure in variants with large cavities. In the V66A/I92A variant an initial increase in intensity of HSQC peaks was observed in the unfolding curves of about 70% of the fitted backbone amides. A similar scenario was observed for about 60% of amides in the L25A/V66A/I92A variant, but not in the I92A/L125A protein. Revisiting data for variant I92A, a localized subgroup of 19 amides shows an increase in intensity, while in variants L125A, V66A, L103A and in the Δ+PHS

protein this behavior was not observed. The spectra were checked for peak overlap that might give rise to this behavior, and suspect peaks were excluded from the analysis. In each protein the increase in intensity of the amides reached a maximum at a given pressure, 800 bar for I92A in 0.65 M GdnHCl, 800 bar for V66A/I92A and 350 bar for L25A/V66A/I92A. This behavior has been previously observed in the OspA protein at high pressure and denoted as a "mountain-like behavior" (57). An increase in the conformational dynamics in the millisecond timescale regime induced by pressure that favors low-lying excited states of lower volume could lead to sharpening of the resonance peaks before the unfolding transition. It is also consistent with a reduction in the dynamics of the native state in the fast (nanosecond) exchange timescale, by which the ensemble of native states in fast exchange with each other is shifted toward states of lower volume (53). This behavior appears to reflect the compression mechanism of the native state and falls in line with the idea that pressure-dependent changes in chemical shifts report on regions of the protein undergoing conformational fluctuations (6, 58). These fluctuations might only involve states within the native ensemble (7), fluctuations across low-lying excited states (59, 60), or local unfolding events (61).

The data for I92A and L125A were collected in 0.65 M and 0.60 M GdnHCl, which could reduce the conformational fluctuations available to the native state and thus highlight the most prominent structural responses to pressure. The plateau value could represent a saturation of the states of lower volume and thus the fully compressed native state or just mark the point in which the unfolding effect outcompetes the phenomena acting on native state fluctuations. It is already known that engineered cavities can modulate the energy landscape of SNase and eliminate an intermediate state (12, 62).

Another proposed origin for the so-called "mountain-like behavior" is the self-association of protein molecules (57). This is unlikely in the case of SNase because one would expect similar associations for Δ+PHS and for its variants bearing substitutions deep in the protein core and far from the exposed surface.

Effect of pressure on the hydration state of internal cavities. The generation of very large cavities in SNase through the use of multiple Ala substitutions raises questions about their state of hydration, despite the crystal structures that suggest they are empty. The linearity between the ΔV measured experimentally and the ΔV calculated from structures suggests that the cavities are empty, but this remains to be verified. An increase in hydrostatic pressure should drive water molecules into the cavities and fill them, thus effectively reducing the measured volume of the cavity. Atomic resolution techniques have provided some insight into the presence of water in protein cavities.

Any putative water molecules that might occupy the internal cavities must be delocalized owing to the lack of internal polar atoms with which to form hydrogen bonds (63). Crystallography is highly insensitive to such dynamic atoms, but applying pressure can localize noble gases within cavities (64) and could affect water in a similar fashion (65). Collins et al. applied pressures of up to 2000 bar on crystals of the cavity containing L99A variant of T4 lysozyme (66). Because spherical, localized density was not detected in the cavity, the authors resorted to integration of electron density within the cavity and assigned the density to up to three buried water molecules. However, the poorly defined density yielded modeled water molecules with very high B-factors (4-5 fold that of the Cβ of Ala-99), making these waters highly questionable. It is difficult to ascertain the

origin of delocalized electron density in a crystallographic experiment, since cosolvent used in the crystallization solution can also localize within internal cavities. It is possible now to download the structure factors and coordinates produced by Collins et al. from RCSB and model in β-mercaptoethanol, a co-solvent used in their experiment, in place of the dubious waters. The cosolvent fits well in the electron density found in the cavity and a single round of refinement results in an improvement of the map and the statistics, making β-mercaptoethanol the clear choice to model over water.

In their NMR spectroscopy experiments, Nucci et al. (67, 68) do not observe peaks corresponding to Nuclear Overhauser Effect (NOE) between ^{13}C and ^{15}N labeled core atoms of ubiquitin and the circa 1000 water molecules confined in their reverse micelle. Fu et al. describe the interior of ubiquitin as dry even when the protein is submitted to pressures of 2500 bar (53). If water molecules are present, they must exchange in sub-nanosecond timescale, the lower limit of detection of NOE. However, when variant L99A of T4 lysozyme was submitted to high pressure in a reverse micelle, Nucci et al. observed new resonances with water molecules involving residues 100 and 101. These amides are close to the cavity generated by substitution L99A and the authors interpreted the waters to reside within the cavity. It is important to note that only the side chain of Ile-100 occludes them from the protein surface, and the water resonances could reflect increased exposure of the amides to bulk solvent. One important conclusion from this study is the fact that increased hydration occurred only at pressures past those required to unfold the protein, but confinement of the protein to a reverse micelle suppressed the unfolding transition. This is inconsistent with pressure driven internal hydration as a mechanism of protein unfolding (20).

A detailed study the role of cavity volume and polarity in determining the hydration state of the protein interior is clearly warranted, focused on structural, dynamic, thermodynamic, and volumetric consequences from water penetration.

Conclusion

A systematic evaluation of the role of cavities as determinants of the pressure unfolding of staphylococcal nuclease has allowed the determination of a correlation between cavity volumes calculated from crystal structures and the ΔV and ΔG of unfolding measured experimentally. Besides highlighting a dominant role for cavities as determinants of the pressure unfolding of proteins, the linear correlation suggests that the total ΔV of SNase is not affected significantly by electrostriction, expansivity and compressibility of the protein, or by the effects of pressure on the properties of water. Site-specific ΔV values obtained using NMR spectroscopy indicate that the region adjacent to cavities are the most sensitive to pressure, displaying the largest responses, perhaps indicative of the largest change in solvent accessible volumes. If the phenomenological relationships between thermodynamic stability and volume changes upon unfolding, and between volume changes upon unfolding and cavity volumes from crystal structures, are found to hold in proteins other than SNase, this will provide a sorely needed framework for subsequent studies of pressure effects on proteins and for the development of parameterized equations for structure-based predictions of pressure effects on proteins. In turn, the site-specific insight gained from high pressure NMR spectroscopy studies coupled with MD simulations provide a method to exploit pressure perturbation and the heterogeneous distribution of void volumes to gain insight into the dynamic nature of the protein interior

Acknowledgements

This work was designed and performed by José Alfredo Caro under the supervision of his advisor Dr. Bertrand García-Moreno E. (Johns Hopkins University, Baltimore, MD). It was possible thanks to a collaboration with Dr. Catherine Ann Royer (Rensselaer Polytechnic Institute, Troy, NY), who served as co-advisor in the design and execution of the experiments, Dr. Jamie L. Schlessman (U.S. Naval Academy, Annapolis, MD) who was instrumental in the production and collection of protein crystals and in the solution of crystal structures, Dr. Mariano Dellarole (Centre de Biochimie Structurale, Montpellier, France), who performed the bulk of the NMR spectroscopy experiments at high pressure and participated in both collection and analysis of fluorescence and NMR spectroscopy data. Dr. Angel E. García (Rensselaer Polytechnic Institute, Troy, NY) is acknowledged for his support and advising in the computational work.

Bibliography

1. Richards FM (1974) The interpretation of protein structures: total volume, group volume distributions and packing density. *J. Mol. Biol.* 82(1):1–14.

2. Chothia C (1975) Structural invariants in protein folding. *Nature* 254(5498):304–308.

3. Richards FM (1977) Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* 6:151–176.

4. Xu J, Baase WA, Baldwin E, Matthews BW (1998) The response of T4 lysozyme to large-to-small substitutions within the core and its relation to the hydrophobic effect. *Protein Sci.* 7(1):158–177.

5. Fleming PJ, Richards FM (2000) Protein packing: dependence on protein size, secondary structure and amino acid composition. *J. Mol. Biol.* 299(2):487–498.

6. Kamatari YO, Smith LJ, Dobson CM, Akasaka K (2011) Cavity hydration as a gateway to unfolding: an NMR study of hen lysozyme at high pressure and low temperature. *Biophys. Chem.* 156(1):24–30.

7. Dellarole M, Roumestand C, Royer CA, Lecomte JTJ (2013) Volumetric properties underlying ligand binding in a monomeric hemoglobin: a high-pressure NMR study. *Biochim. Biophys. Acta* 1834(9):1910–22.

8. Lee C, Maeng J, Kocher J, Lee B (2001) Cavities of α1-antitrypsin that play structural and functional roles. *Protein Sci.* 10(7):1446–1453.

9. Wiener R, Zhang X, Wang T, Wolberger C (2012) The mechanism of OTUB1-mediated inhibition of ubiquitination. *Nature* 483(7391):618–622.

10. Tokuriki N, Tawfik DS (2009) Protein Dynamism and Evolvability. *Science* 324(5924):203–207.

11. López CJ, Yang Z, Altenbach C, Hubbell WL (2013) Conformational selection and adaptation to ligand binding in T4 lysozyme cavity mutants. *Proc. Natl. Acad. Sci. U.S.A.* 110(46):E4306–E4315.

12. Roche J, Caro JA, Norberto DR, Barthe P, Roumestand C, Schlessman JL, Garcia AE, García-Moreno E. B, Royer CA (2012) Cavities determine the pressure unfolding of proteins. *Proc. Natl. Acad. Sci. U.S.A.* 109(18):6945–6950.

13. Ando N, Barstow B, Baase WA, Fields A, Matthews BW, Gruner SM (2008) Structural and thermodynamic characterization of T4 lysozyme mutants and the

contribution of internal cavities to pressure denaturation. *Biochemistry* 47(42):11097–11109.

14. Urbauer JL, Ehrhardt MR, Bieber RJ, Flynn PF, Wand AJ (1996) High-Resolution Triple-Resonance NMR Spectroscopy of a Novel Calmodulin-Peptide Complex at Kilobar Pressures. *J. Am. Chem. Soc.* 118(45):11329–11330.

15. Chalikian TV, Macgregor Jr RB (2009) Origins of pressure-induced protein transitions. *J. Mol. Biol.* 394(5):834–842.

16. Vasilchuk D, Pandharipande PP, Suladze S, Sanchez-Ruiz JM, Makhatadze GI (2014) Molecular determinants of expansivity of native globular proteins: a pressure perturbation calorimetry study. *J. Phys. Chem. B* 118(23):6117–6122.

17. Dellarole M, Kobayashi K, Rouget J-B, Caro JA, Roche J, Islam MM, García-Moreno E. B, Kuroda Y, Royer CA (2013) Probing the Physical Determinants of Thermal Expansion of Folded Proteins. *J. Phys. Chem. B* 117(42):12742–12749.

18. Frye KJ, Perman CS, Royer CA (1996) Testing the correlation between ΔA and ΔV of protein unfolding using m value mutants of staphylococcal nuclease. *Biochemistry* 35(31):10234–10239.

19. Rouget J-B, Aksel T, Roche J, Saldana J-L, Garcia AE, Barrick D, Royer CA (2011) Size and sequence and the volume change of protein folding. *J. Am. Chem. Soc.* 133(15):6020–6027.

20. Hummer G, Garde S, García AE, Paulaitis ME, Pratt LR (1998) The pressure dependence of hydrophobic interactions is consistent with the observed pressure denaturation of proteins. *Proc. Natl. Acad. Sci. U.S.A.* 95(4):1552–1555.

21. Prehoda KE, Mooberry ES, Markley JL (1998) Pressure denaturation of proteins: evaluation of compressibility effects. *Biochemistry* 37(17):5785–5790.

22. Kharakoz DP (2000) Protein compressibility, dynamics, and pressure. *Biophys. J.* 79(1):511–525.

23. Lin L-N, Brandts JF, Brandts JM, Plotnikov V (2002) Determination of the volumetric properties of proteins and other solutes using pressure perturbation calorimetry. *Anal. Biochem.* 302(1):144–160.

24. Seemann H, Winter R, Royer CA (2001) Volume, expansivity and isothermal compressibility changes associated with temperature and pressure unfolding of Staphylococcal nuclease. *J. Mol. Biol.* 307(4):1091–1102.

25. Mitra L, Smolin N, Ravindra R, Royer CA, Winter R (2006) Pressure perturbation calorimetric studies of the solvation properties and the thermal unfolding of

proteins in solution-experiments and theoretical interpretation. *Phys. Chem. Chem. Phys.* 8(11):1249–1265.

26. Wiedersich J, Köhler S, Skerra A, Friedrich J (2008) Temperature and pressure dependence of protein stability: the engineered fluorescein-binding lipocalin FluA shows an elliptic phase diagram. *Proc. Natl. Acad. Sci. U.S.A.* 105(15):5756–61.

27. Vajpai N, Nisius L, Wiktor M, Grzesiek S (2013) High-pressure NMR reveals close similarity between cold and alcohol protein denaturation in ubiquitin. *Proc. Natl. Acad. Sci. U.S.A.* 110(5):E368–E376.

28. Herberhold H, Royer CA, Winter R (2004) Effects of chaotropic and kosmotropic cosolvents on the pressure-induced unfolding and denaturation of proteins: An FT-IR study on staphylococcal nuclease. *Biochemistry* 43(12):3336–3345.

29. Wang S, Tate MW, Gruner SM (2012) Protein crowding impedes pressure-induced unfolding of staphylococcal nuclease. *Biochim. Biophys. Acta* 1820(7):957–961.

30. Ravindra R, Winter R (2004) Pressure perturbation calorimetry: a new technique provides surprising results on the effects of co-solvents on protein solvation and unfolding behaviour. *Chemphyschem* 5(4):566–571.

31. Ravindra R, Royer CA, Winter R (2004) Pressure perturbation calorimetic studies of the solvation properties and the thermal unfolding of staphylococcal nuclease. *Phys. Chem. Chem. Phys.* 6(8):1952–1961.

32. Grigera JR, McCarthy AN (2010) The behavior of the hydrophobic effect under pressure and protein denaturation. *Biophys. J.* 98(8):1626–1631.

33. Kabsch W (1976) A solution for best rotation to relate 2 sets of vectors. *Acta Crystallogr. Sect. A* 32(5):922–923.

34. Kleywegt GJ, Jones TA (1994) Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr. Sect. D - Biol. Crystallogr.* 50(2):178–185.

35. Till MS, Ullmann GM (2010) McVol - a program for calculating protein volumes and identifying cavities by a Monte Carlo algorithm. *J. Mol. Model.* 16(3):419–429.

36. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* 34(suppl 2):W116–W118.

37. Damjanović A, García-Moreno E. B, Lattman EE, García AE (2005) Molecular dynamics study of water penetration in staphylococcal nuclease. *Proteins Struct. Funct. Bioinforma.* 60(3):433–49.

38. Damjanović A, Schlessman JL, Fitch CA, García AE, García-Moreno E. B (2007) Role of flexibility and polarity as determinants of the hydration of internal cavities and pockets in proteins. *Biophys. J.* 93(8):2791–804.

39. Denisov VP, Schlessman JL, García-Moreno E. B, Halle B (2004) Stabilization of internal charges in a protein: water penetration or conformational change? *Biophys. J.* 87(December):3982–3994.

40. Isom DG, Castañeda CA, Cannon BR, García-Moreno E. B (2011) Large shifts in pK a values of lysine residues buried inside a protein. *Proc. Natl. Acad. Sci. U.S.A.* 108(13):5260–5265.

41. Eriksson AE, Baase WA, Zhang XJ, Heinz DW, Blaber M, Baldwin EP, Matthews BW (1992) Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* 255(5041):178–183.

42. Llinás M, Gillespie B, Dahlquist FW, Marqusee S (1999) The energetics of T4 lysozyme reveal a hierarchy of conformations. *Nat. Struct. Biol.* 6(11):1072–1078.

43. Llinás M, Marqusee S (1998) Subdomain interactions as a determinant in the folding and stability of T4 lysozyme. *Protein Sci.* 7(1):96–104.

44. Cellitti J, Llinas M, Echols N, Shank E a, Gillespie B, Kwon E, Crowder SM, Dahlquist FW, Alber T, Marqusee S (2007) Exploring subdomain cooperativity in T4 lysozyme I: structural and energetic studies of a circular permutant and protein fragment. *Protein Sci.* 16(5):842–851.

45. Cellitti J, Bernstein R, Marqusee S (2007) Exploring subdomain cooperativity in T4 lysozyme II: uncovering the C-terminal subdomain as a hidden intermediate in the kinetic folding pathway. *Protein Sci.* 16(5):852–862.

46. Spencer D, García-Moreno E. B, Stites WE (2013) The pH dependence of staphylococcal nuclease stability is incompatible with a three-state denaturation model. *Biophys. Chem.* 180-181:86–94.

47. Shortle D (1995) Staphylococcal nuclease: a showcase of m-value effects. *Adv. Protein Chem.* 46:217–247.

48. Shortle D, Stites WE, Meeker AK (1990) Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease. *Biochemistry* 29(35):8033–41.

49. Toptygin D, Woolf TB, Brand L (2010) Picosecond protein dynamics: the origin of the time-dependent spectral shift in the fluorescence of the single Trp in the protein GB1. *J. Phys. Chem. B* 114(34):11323–37.

50. Sturtevant JM (1987) Biochemical applications of differential scanning calorimetry. *Annu. Rev. Phys. Chem.* 38(1):463–488.

51. Spencer D, Bertrand G-ME, Stites WE (2013) The pH dependence of staphylococcal nuclease stability is incompatible with a three-state denaturation model. *Biophys. Chem.* 180-181:86–94.

52. Eriksson AE, Baase WA, Matthews BW (1992) Similar hydrophobic replacements of Leu99 and Phe153 within the core of T4 lysozyme have different structural and thermodynamic consequences. *J. Mol. Biol.* 229(3):747–769.

53. Fu Y, Kasinath V, Moorman VR, Nucci NV, Hilser VJ, Wand AJ (2012) Coupled motion in proteins revealed by pressure perturbation. *J. Am. Chem. Soc.* 134(20):8543–8550.

54. Fuentes EJ, Gilmore SA, Mauldin RV, Lee AL (2006) Evaluation of energetic and dynamic coupling networks in a PDZ domain protein. *J. Mol. Biol.* 364(3):337–51.

55. Fuentes EJ, Der CJ, Lee AL (2004) Ligand-dependent Dynamics and Intramolecular Signaling in a PDZ Domain. *J. Mol. Biol.* 335(4):1105–1115.

56. Damjanović A, García-Moreno E. B, Lattman EE, García AE (2005) Molecular dynamics study of hydration of the protein interior. *Comput. Phys. Commun.* 169(1-3):126–129.

57. Kitahara R, Simorellis AK, Hata K, Maeno A, Yokoyama S, Koide S, Akasaka K (2012) A delicate interplay of structure, dynamics, and thermodynamics for function: a high pressure NMR study of outer surface protein A. *Biophys. J.* 102(4):916–926.

58. Akasaka K (2003) Highly Fluctuating Protein Structures Revealed by Variable-Pressure Nuclear Magnetic Resonance. *Biochemistry* 42(37):10875–10885.

59. Akasaka K, Li H (2001) Low-Lying Excited States of Proteins Revealed from Nonlinear Pressure Shifts in 1H and 15N NMR. *Biochemistry* 40(30):8665–8671.

60. Maeno A, Sindhikara D, Hirata F, Otten R, Frederick W, Yokoyama S, Akasaka K, Mulder F (2014) Cavity as a source of conformational fluctuation and high-energy state : High-pressure NMR study of a cavity-enlarged mutant of T4 lysozyme. *Biophys. J.*

61. Nucci NV, Fuglestad B, Athanasoula EA, Wand AJ (2014) Role of cavities and hydration in the pressure unfolding of T4 lysozyme. *Proc. Natl. Acad. Sci. U.S.A.* 111(38):13846–13851.

62. Roche J, Dellarole M, Caro JA, Norberto DR, Garcia AE, García-Moreno E. B, Roumestand C, Royer CA (2013) Effect of Internal Cavities on Folding Rates and Routes Revealed by Real-time Pressure-Jump NMR Spectroscopy. *J. Am. Chem. Soc.* 82(6):1069–1080.

63. Schlessman JL, Abe C, Gittis A, Karp DA, Dolan MA, García-Moreno E. B (2008) Crystallographic study of hydration of an internal cavity in engineered proteins with buried polar or ionizable groups. *Biophys. J.* 94(8):3208–16.

64. Quillin ML, Breyer W a, Griswold IJ, Matthews BW (2000) Size versus polarizability in protein-ligand interactions: binding of noble gases within engineered cavities in phage T4 lysozyme. *J. Mol. Biol.* 302(4):955–977.

65. Nagae T, Kawamura T, Chavas LMG, Niwa K, Hasegawa M, Kato C, Watanabe N (2012) High-pressure-induced water penetration into 3-isopropylmalate dehydrogenase. *Acta Crystallogr. Sect. D - Biol. Crystallogr.* 68(3):300–309.

66. Collins MD, Hummer G, Quillin ML, Matthews BW, Gruner SM (2005) Cooperative water filling of a nonpolar protein cavity observed by high-pressure crystallography and simulation. *Proc. Natl. Acad. Sci. U.S.A.* 102(46):16668–16671.

67. Nucci NV, Pometun MS, Wand AJ (2011) Site-resolved measurement of water-protein interactions by solution NMR. *Nat. Struct. Mol. Biol.* 18(2):245–249.

68. Nucci NV, Pometun MS, Wand AJ (2011) Mapping the hydration dynamics of ubiquitin. *J. Am. Chem. Soc.* 133(32):12326–12329.

69. The PyMOL Molecular Graphics System, Version 1.3 Schrödinger, LLC.

Tables

Table 1. Thermodynamic stability from chemical denaturation [a, b]

Protein	ΔG^{o}_{H2O} (kcal/mol)	$\Delta\Delta G^{o}_{H2O}$ (kcal/mol)	$\Delta\Delta G^{o}_{Expected}$ (kcal/mol)	$\Delta\Delta G^{o}_{Measured} - \Delta\Delta G^{o}_{Expected}$ (kcal/mol)	*m*-value (kcal/mol*M)	C_{mid} (M)
Δ+PHS	11.9 (0.1)	--	--	--	4.9 (0.1)	2.4 (0.1)
V23A	8.7 (0.1)	-3.2 (0.1)	--	--	5.2 (0.1)	1.7 (0.1)
L25A	8.9 (0.1)	-3.0 (0.1)	--	--	5.0 (0.1)	1.8 (0.1)
F34A	8.3 (0.1)	-3.6 (0.1)	--	--	5.2 (0.1)	1.6 (0.1)
L36A	8.6 (0.1)	-3.3 (0.1)	--	--	5.4 (0.1)	1.6 (0.1)
L38A	10.5 (0.1)	-1.4 (0.1)	--	--	4.8 (0.1)	2.2 (0.1)
V66A	9.0 (0.2)	-2.9 (0.2)	--	--	4.8 (0.1)	1.9 (0.2)
V74A	8.5 (0.2)	-3.4 (0.2)	--	--	5.0 (0.1)	1.7 (0.1)
I92A	7.9 (0.3)	-4.0 (0.3)	--	--	5.4 (0.2)	1.5 (0.2)
L103A	7.8 (0.1)	-4.1 (0.1)	--	--	5.2 (0.1)	1.5 (0.1)
L125A	8.1 (0.1)	-3.8 (0.1)	--	--	5.5 (0.1)	1.5 (0.1)
V23A/I92A	5.1 (0.1)	-6.8 (0.1)	-7.2	0.4	5.9 (0.1)	0.9 (0.1)
L25A/I92A	5.4 (0.1)	-6.5 (0.1)	-7.0	0.5	5.6 (0.1)	1.0 (0.1)
L36A/I92A	2.5 (0.1)	-9.4 (0.1)	-7.3	-2.1	6.3 (0.1)	0.4 (0.1)
V66A/I92A	5.6 (0.2)	-6.3 (0.2)	-6.9	0.6	5.4 (0.2)	1.0 (0.1)
V74A/I92A	4.0 (0.1)	-7.9 (0.1)	-7.4	-0.5	5.9 (0.1)	0.7 (0.1)
I92A/L103A	3.7 (0.1)	-8.2 (0.1)	-8.1	-0.1	6.0 (0.2)	0.6 (0.1)
I92A/L125A	3.6 (0.1)	-8.3 (0.1)	-7.8	-0.5	6.0 (0.1)	0.6 (0.1)
L25A/V66A/I92A	3.5 (0.1)	-8.4 (0.1)	-9.9	1.5	6.0 (0.1)	0.6 (0.1)
L25A/L36A/I92A	0.4 (0.1)	-11.5 (0.1)	-10.3	-1.2	6.4 (0.1)	0.0 (0.1)

[a]All experiments were performed in 25 mM HEPES with 100 mM KCl at 25 °C, pH 7

[b]Values in parenthesis represent the error of the fit. A single experiment was performed for each protein.

Table 2. Thermodynamic parameters from pressure unfolding experiments monitored by fluorescence.

Protein	Conditions (M Gdn)	$\Delta G^{o}_{(1\ bar)}$ (kcal/mol)	ΔV (mL/mol)	P_{mid} (bar)
Δ+PHS	2	1.8 (0.1)	29 (3)	2597 (1460)
	2.3	0.8 (0.1)	58 (4)	577 (310)
V23A	0.8	4.0 (0.2)	74 (4)	2262 (1076)
	1	3.3 (0.3)	74 (6)	1866 (1337)
	1.4	1.6 (0.2)	62 (7)	1080 (975)
L25A	1	3.8 (0.1)	89 (3)	1786 (575)
	1.2	3.2 (0.1)	96 (4)	1395 (575)
	1.4	2.1 (0.2)	87 (7)	1010 (771)
F34A	0.8	3.7 (0.1)	73 (3)	2121 (753)
	1	3.0 (0.1)	83 (2)	1512 (343)
	1.2	2.1 (0.1)	74 (3)	1187 (422)
L36A	0.9	3.4 (0.2)	95 (4)	1497 (635)
	1.1	2.8 (0.1)	104 (4)	1126 (447)
	1.3	1.9 (0.2)	101 (7)	787 (560)
L38A	1.6	2.7 (0.1)	68 (3)	1661 (613)
	1.8	2.2 (0.1)	70 (4)	1315 (635)
	2	1.5 (0.2)	79 (6)	794 (551)
V66A	1.2	4.1 (0.1)	87 (3)	1972 (642)
	1.4	2.8 (0.2)	70 (4)	1674 (825)
	1.6	1.8 (0.1)	66 (3)	1141 (430)
V74A	1	3.2 (0.1)	69 (2)	1940 (480)
	1.2	2.3 (0.1)	65 (1)	1480 (208)
	1.4	1.8 (0.1)	81 (3)	930 (318)
I92A	0.8	3.4 (0.2)	110 (6)	1293 (753)
	1	2.8 (0.1)	113 (4)	1037 (395)
	1.2	2.1 (0.1)	130 (7)	676 (418)
L103A	0.8	3.0 (0.1)	72 (2)	1743 (423)
	1	2.3 (0.2)	79 (5)	1218 (704)
	1.4	1.0 (0.2)	90 (10)	465 (499)
L125A	0.8	3.3 (0.2)	80 (4)	1726 (795)
	1	2.5 (0.1)	81 (3)	1291 (438)
	1.2	1.8 (0.1)	74 (4)	1018 (479)

Table 2. continued.

Protein	Conditions (M Gdn)	$\Delta G^{o}_{(1\ bar)}$ (kcal/mol)	ΔV (mL/mol)	P_{mid} (bar)
V23A/I92A	0	4.4 (0.2)	107 (4)	1721 (685)
	0.35	3.0 (0.1)	129 (3)	973 (263)
L25A/I92A	0	4.8 (0.2)	116 (5)	1731 (819)
	0.35	3.5 (0.1)	135 (5)	1085 (470)
L36A/I92A	0	2.7 (0.1)	136 (5)	831 (360)
V66A/I92A	0	5.7 (0.2)	124 (3)	1923 (543)
	0.4	3.5 (0.1)	118 (5)	1235 (572)
V74A/I92A	0	3.8 (0.2)	130 (6)	1223 (656)
I92A/L103A	0	4.0 (0.2)	133 (5)	1268 (564)
I92A/L125A	0	3.7 (0.1)	131 (3)	1182 (316)
L25A/V66A/I92A	0	3.4 (0.1)	147 (6)	968 (482)
L25A/L36A/I92A	0	0.6 (0.1)*	170 (4)	3 (1)

[a]All experiments were performed in 50 mM Tris at 20 °C, pH 7.

Table 3. Thermodynamic parameters from pressure unfolding experiments monitored by NMR spectroscopy.

Protein	No. intensity profiles fitted	Conditions (M Gdn)	$\Delta G^{o}_{(1bar)}$ (kcal/mol)	$\Delta V_{average}$ (mL/mol)	P_{mid} (bar)
Δ+PHS	88	1.8	2.7 (0.4)	89 (13)	1269
V66A	107	1.1	3.7 (1.1)	114 (33)	1358
I92A	89	0.65	4.4 (0.8)	130 (21)	1416
L125A	93	0.85	3.2 (0.3)	116 (10)	1154
V66A/I92A	98	0	5.7 (1.0)	144 (25)	1656
I92A/L125A	85	0	4.4 (0.5)	192 (22)	959
L25A/V66A/I92A	83	0	4.0 (0.6)	201 (27)	833

Figure legends

Figure 1: Crystal structures of Δ+PHS SNase and of Ala-substituted variants. Cavities are shown in red mesh calculated using a molecular surface in PyMOL (69) and a probe of 1.4 Å. PDB accession numbers for these proteins are listed in Table 1. **(A)** Δ+PHS, **(B)** V23A, **(C)** L25A, **(D)** F34A, **(E)** L36A, **(F)** L38A, **(G)** V66A, **(H)** V74A, **(I)** I92A, **(J)** L103A, **(K)** L125A, **(L)** V23A/I92A, **(M)** L25A/I92A, **(N)** V66A/I92A, **(O)** V74A/I92A, **(P)** I92A/L103A, **(Q)** I92A/L125A, **(R)** L25A/L36A/I92A, **(S)** L25A/V66A/I92A. The variants with a single alanine substitution labeled A-K have been published previously (12).

Figure 2: (A) Chemical denaturation monitored by Trp fluorescence for variants with I92A, V23A/I92A, L25A/I92A, L36A/I92A, V66A/I92A, V74A/I92A, I92A/L103A, I92A/L125A, L25A/L36A/I92A, and L25A/V66A/I92A. Δ+PHS is highlighted in bold. **(B)** Pressure unfolding monitored with Trp fluorescence. GdnHCl was used for variants V23A/I92A (0.35M), L25A/I92A (0.35M) and V66A/I92A (0.4M), as shown on Table 2. The pressure unfolding curve of the variant I92A in 1.0M GdnHCl is highlighted in bold.

Figure 3: Intensity profiles of HSQC cross-peaks for variants with the following substitutions: **(A)** I92A, **(B)** V66A/I92A, **(C)** I92A/L125A, and **(D)** L25A/V66A/I92A. The profiles were normalized individually to the value measured at 1 bar. The dotted lines represent continuously decaying profiles that were not fitted. Histograms of ΔV values obtained from HSQC intensity profiles for variants with the following substitutions: **(E)** I92A, **(F)** V66A/I92A, **(G)** I92A/L125A, and **(H)** L25A/V66A/I92A.

Gaussian fits of the histogram result in the following average values (standard deviation): V66A/I92A, <ΔVnmr> = 144 (±18) mL/mol; I92A/L125A, <ΔVnmr> = 189 (±21) mL/mol; L25A/V66A/I92A, <ΔVnmr> = 199 (±29) mL/mol.

Figure 4: Mapping of ΔV onto the structures of **(A)** V66A/I92A, **(B)** I92A/L125A, and **(C)** L25A/V66A/I92A, ranging from smallest (yellow) to largest (red) values. White represents amides that were not fitted. **(D-F)** Structural mapping of ΔΔV ($\Delta V_{variant}$ - $\Delta V_{background}$) for the proteins shown in **(A-C)**, respectively. Red identifies the 25% of amides with the largest ΔΔV. Yellow identifies all other ΔΔV >0. Blue identifies ΔΔV < 0.

Figure 5: ΔV values for Δ+PHS and the 19 variants with one, two or three Ala substitutions. Black, gray and light gray bars represent values obtained from fluorescence experiments, with the light gray values corresponding to work previously published and black representing Δ+PHS, the parent protein (12). Filled squares (■) are the average ΔV values and standard deviations obtained by NMR spectroscopy. Crosses (×) are the values predicted by additivity, obtained by adding the effects on ΔV of the individual substitutions.

Figure 6: Correlation plot of ΔV measured with pressure unfolding experiments monitored with Trp fluorescence and ΔV calculated from cavities using CASTp and probe of **(A)** 1.4 Å or **(B)** 1.2 Å or with **(C)** difference in van der Waals volume of side chains (3). (□) identifies, Δ+PHS (○) identifies variants with single Ala substitutions

and (●) are the variants with two or three substitutions to Ala. The linear fit is represented by the solid line. The dotted line represents a line with unity slope.

Figure 7: Correlation between ΔG^{o}_{H2O} measured by GdnHCl titrations and ΔV measured from pressure unfolding experiments monitored with Trp fluorescence. (□) identifies Δ+PHS, (○) identifies variants with single Ala substitutions and (●) are the variants with two or three substitutions to Ala. The slope and intercept of the line are 84 ± 8 cal/mL and -3.7 ± 0.9 kcal/mol.

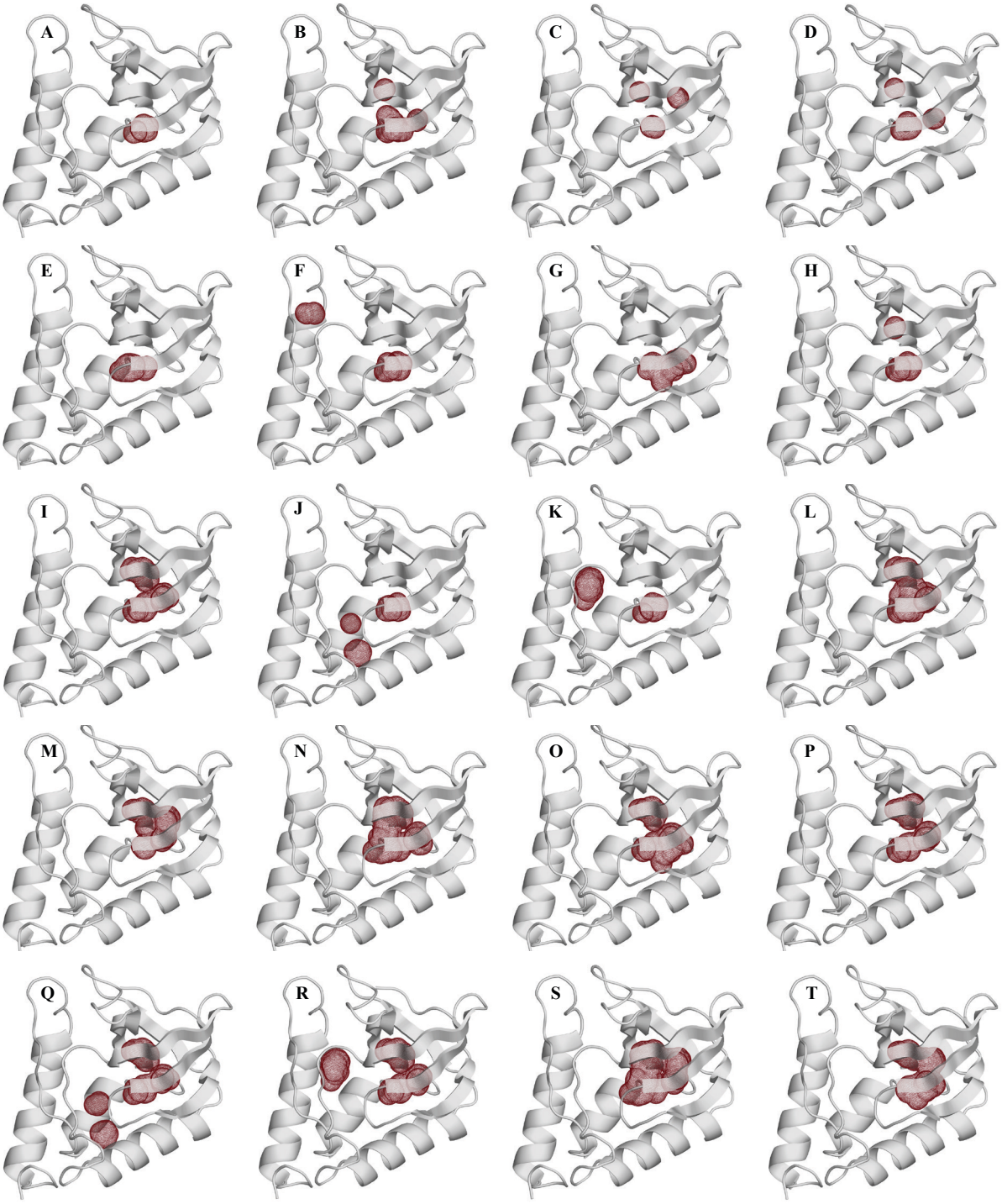


Figure 1

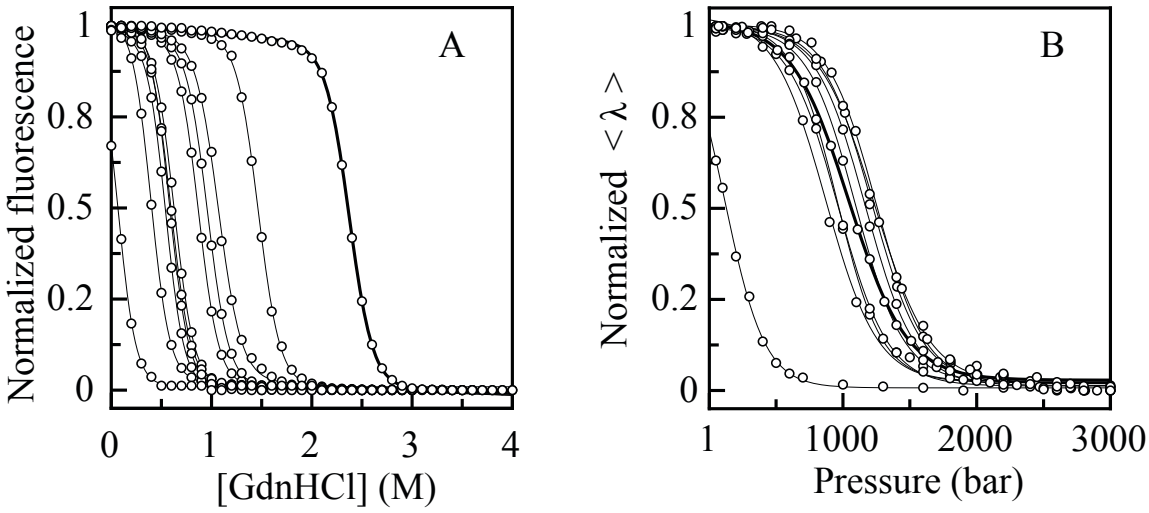

A
B
Normalized fluorescence
Normalized < λ >
[GdnHCl] (M)
Pressure (bar)
0
0.2
0.5
0.8
1
0
1
2
3
4
1
1000
2000
3000

Figure 2

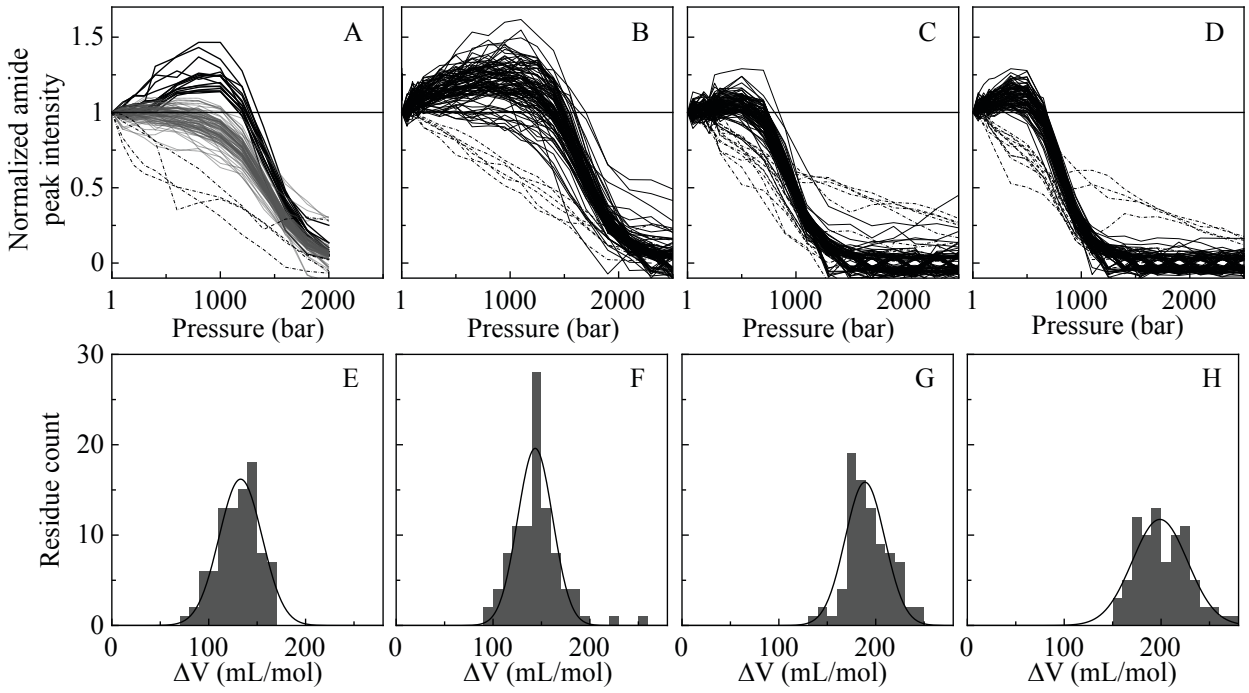

A
B
C
D
Normalized amide peak intensity
1.5
1
0.5
0
1
1000
2000
Pressure (bar)
E
F
G
H
Residue count
30
20
10
0
100
200
ΔV (mL/mol)

Figure 3

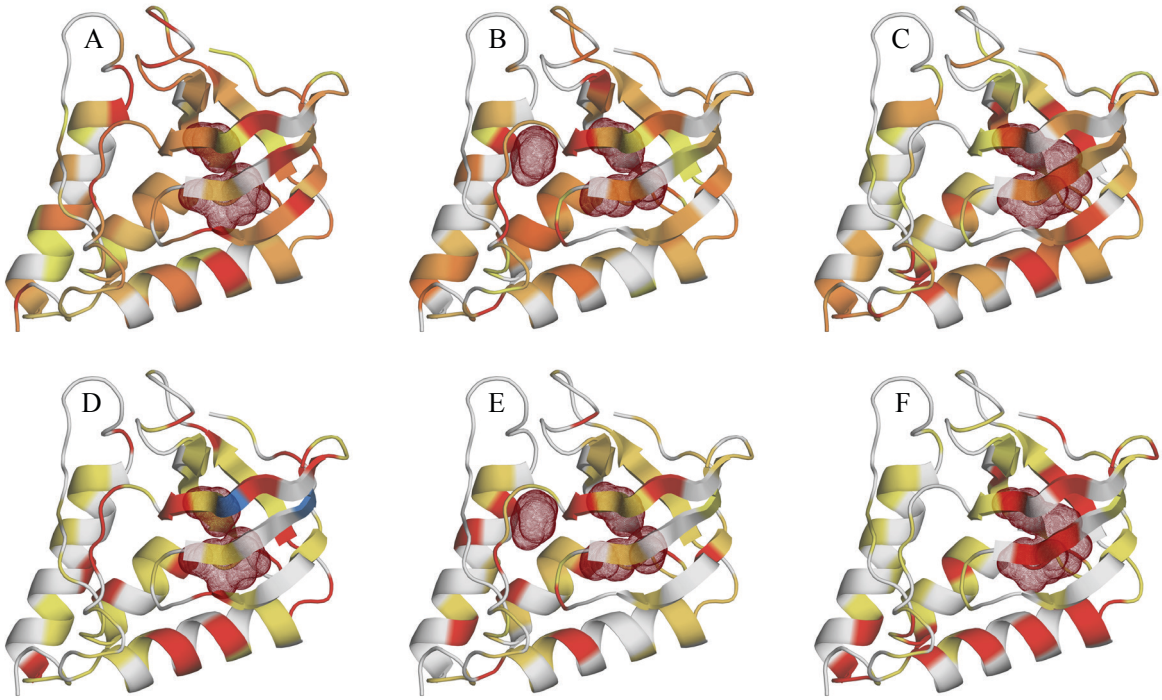


Figure 4

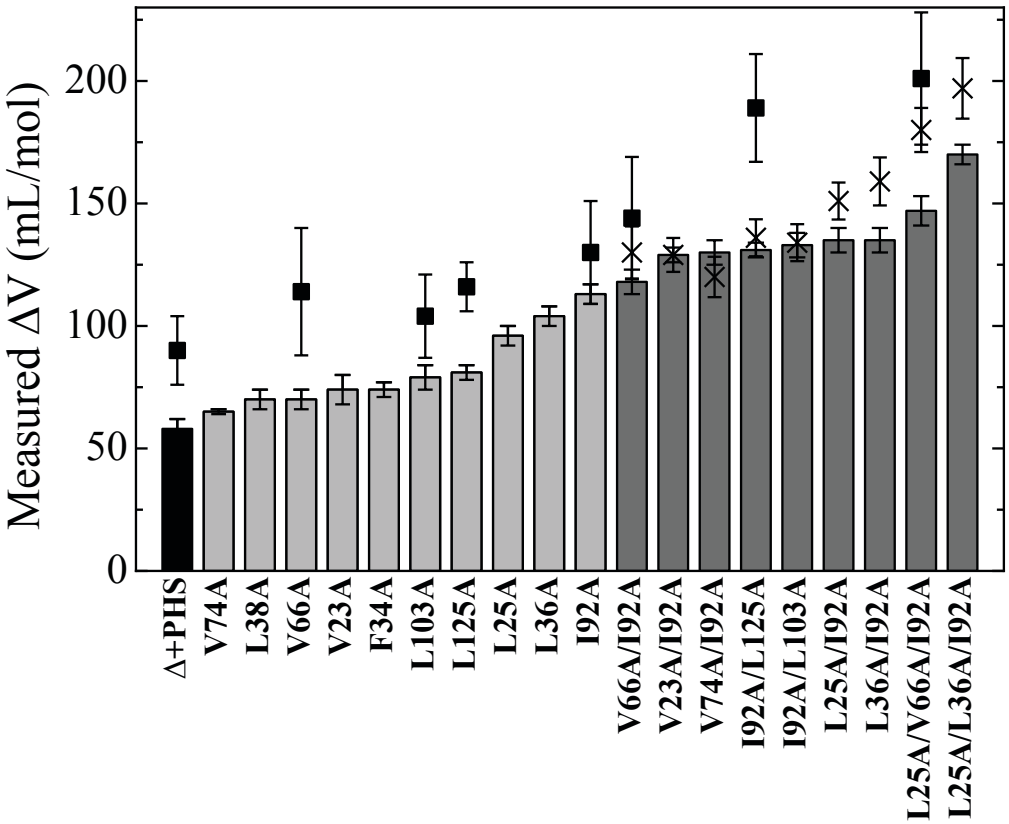

Measured ΔV (mL/mol)
0
50
100
150
200
Δ+PHS
V74A
L38A
V66A
V23A
F34A
L103A
L125A
L25A
L36A
I92A
V66A/I92A
V23A/I92A
V74A/I92A
I92A/L125A
I92A/L103A
L25A/I92A
L36A/I92A
L25A/V66A/I92A
L25A/L36A/I92A

Figure 5

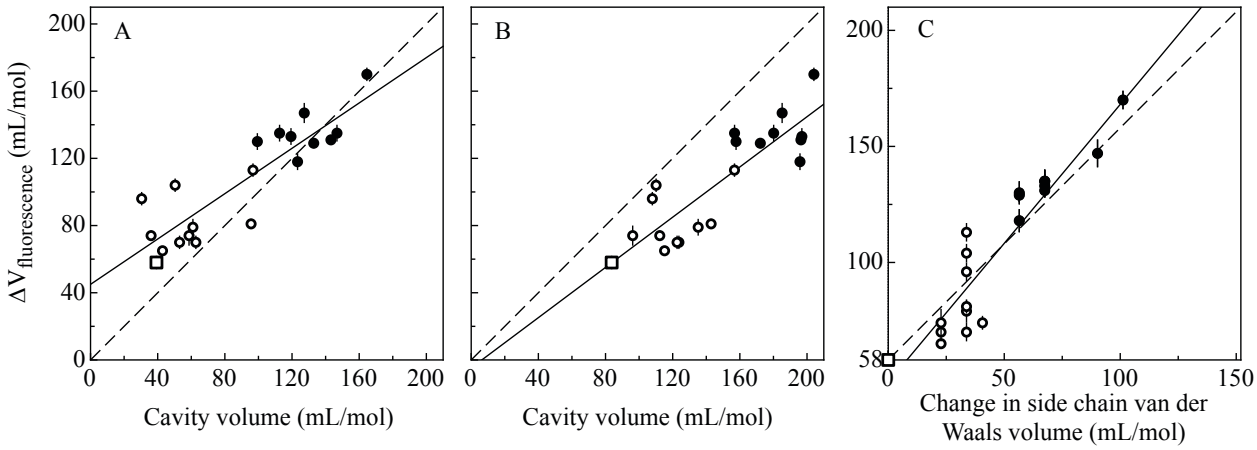

A
B
C
$\Delta V_{fluorescence}$ (mL/mol)
Cavity volume (mL/mol)
Cavity volume (mL/mol)
Change in side chain van der Waals volume (mL/mol)

Figure 6

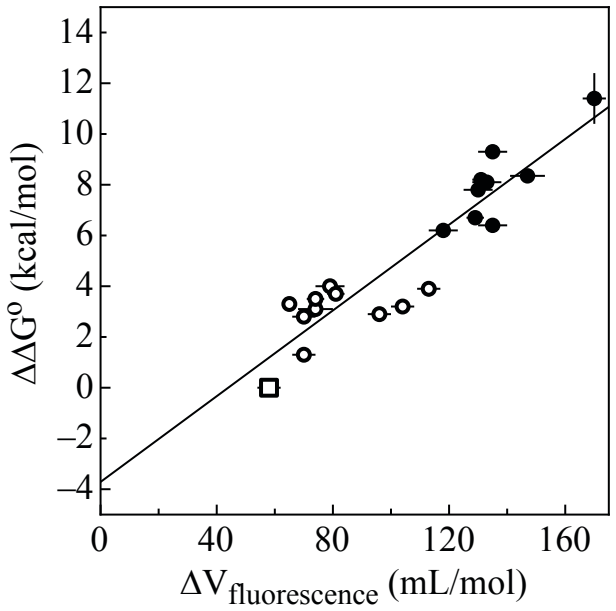

14
12
10
8
6
4
2
0
–2
–4
0
40
80
120
160
ΔΔG° (kcal/mol)
ΔV fluorescence (mL/mol)

Figure 7

Table S1. Volume calculations from crystal structures.

Protein	Change in van der Waals volume (mL/mol)	Probe radius (Å)	Calculated volume (mL/mol)			PDB ID
			McVol	CASTp	VOIDOO	
Δ+PHS	0.00	1	--	--	29 (17)	3BDC
		1.2	40	84	--	
		1.4	11	39	0	
V23A	22.76	1	--	--	50 (13)	3PMF
		1.2	60	96	--	
		1.4	33	59	0	
L25A	33.73	1	--	--	17 (7)	3OSO
		1.2	58	108	--	
		1.4	7	30	0	
F34A	40.62	1	--	--	41 (31)	3MVV
		1.2	41	112	--	
		1.4	16	36	0	
L36A	33.73	1	--	--	42 (17)	3NP8
		1.2	47	110	--	
		1.4	20	50	0	
L38A	33.73	1	--	--	60 (27)	3MHB
		1.2	64	124	--	
		1.4	30	53	0	
V66A	22.76	1	--	--	55 (32)	3NQT
		1.2	92	123	--	
		1.4	30	63	0	
V74A	22.76	1	--	--	50 (19)	3NK9
		1.2	52	115	--	
		1.4	21	43	0	
I92A	33.73	1	--	--	117 (63)	3MEH
		1.2	87	157	--	
		1.4	56	97	0	
L103A	33.73	1	--	--	46 (19)	3MZ5
		1.2	38	135	--	
		1.4	42	61	0	
L125A	33.73	1	--	--	52 (15)	3NXW
		1.2	87	143	--	
		1.4	58	96	12 (7)	

Table S1 (continued)

Protein	Change in van der Waals volume (mL/mol)	Probe radius (Å)	Calculated volume (mL/mol)			PDB ID
			McVol	CASTp	VOIDOO	
V23A/I92A	56.49	1	--	--	184 (74)	3VA5
		1.2	113	172	--	
		1.4	91	133	67 (25)	
L25A/I92A	67.46	1	--	--	194 (71)	4F7X
		1.2	102	157	--	
		1.4	49	113	20 (7)	
L36A/I92A	67.46	1	--	--	133 (64)	4DFA
		1.2	135	180	--	
		1.4	105	148	45 (12)	
		1.2 (MD)	139 (19)	--	--	
V66A/I92A	56.49	1	--	--	159 (61)	3V2T
		1.2	117	196	--	
		1.4	76	123	30 (7)	
V74A/I92A	56.49	1	--	--	50 (19)	4DU9
		1.2	85	158	--	
		1.4	56	99	0	
I92A/L103A	67.46	1	--	--	90 (50)	3SXH
		1.2	112	197	--	
		1.4	71	119	0	
I92A/L125A	67.46	1	--	--	100 (41)	4DGZ
		1.2	131	196	--	
		1.4	94	143	13 (5)	
L25A/L36A/I92A	101.19	1	--	--	268 (87)	4K2K
		1.2	150	204	--	
		1.4	103	165	84 (23)	
		1.2 (MD)	152 (19)	--	--	
L25A/V66A/I92A	90.22	1	--	--	262 (134)	4KD3
		1.2	121	185	--	
		1.4	61	127	21 (7)	

-- Values under probe size 1.2 (MD) were obtained from 10 ns molecular dynamics simulations in combination with McVol and a probe of 1.2 Å.

Table S2. Crystallography tables.

Protein	Δ+PHS V23A/I92A	Δ+PHS L25A/I92A
Crystallization conditions		
Buffer	25mM K phosphate	25mM K phosphate
pH	7	6
Temperature (K)	277	277
Precipitant	20% (w/v) MPD	20% (w/v) MPD
Additives	pdTp, $CaCl_2$	pdTp, $CaCl_2$
Data collection		
Wavelength (Å)	0.97950	1.1
Resolution (Å)	50.00-1.55 (1.58-1.55)	50.00-1.70 (1.73-1.70)
Unique reflections	20327 (865)	15561 (783)
Completeness	0.982 (0.859)	0.998 (0.999)
Redundancy	5.8 (3.5)	6.1 (5.5)
Average $I/\sigma(I)$	11.6 (3.5)	13.5 (10.3)
R_{merge}	0.066 (0.336)	0.071 (0.167)
Wilson B ($Å^2$)	32.3	30.7
Space group	$P2_1$	$P2_1$
Cell dimensions (Å ; °)	a = 31.18 ; α = 90.00 b = 60.27 ; β = 94.78 c = 38.10 ; γ = 90.00	a = 31.12 ; α = 90.00 b = 59.91 ; β = 95.03 c = 38.48 ; γ = 90.00
Refinement		
Resolution (Å)	32.12-1.55 (1.59-1.55)	38.33-1.70 (1.74-1.70)
No. of non-hydrogen atoms	1204	1227
No. of unique reflections	20146 (1346)	15543 (1076)
No. of reflections in test set	2046 (138)	776 (57)
R_{work}	0.175 (0.31)	0.163 (0.19)
R_{free}	0.219 (0.29)	0.197 (0.22)
RMS from ideal geometry		
Bonds (Å)	0.017	0.016
RMS angles (°)	1.72	1.87
Average B-factors ($Å^2$)		
Protein	25.9	23.5
Solvent	32.5	30.5
Ion	24.2	19.5
Ramachandran plot		
Most favored (%)	100 (87.7)	99 (86.8)
Additionally allowed (%)	13 (11.4)	14 (12.3)
Disallowed (%)	1 (0.9)	1 (0.9)
No. of residues excluding Gly, Pro and termini	114	114
Total no. of residues	129	129
PDB accession code	3VA5	4F7X
RMSD (Å) from Δ+PHS	0.33	0.37
Main chain only	0.15	0.19

Protein	Δ+PHS L36A/I92A	Δ+PHS I92A/V66A
Crystallization conditions		
Buffer	25mM K phosphate	25mM K phosphate
pH	7	6
Temperature (K)	277	277
Precipitant	20% (w/v) MPD	20% (w/v) MPD
Additives	pdTp, $CaCl_2$	pdTp, $CaCl_2$
Data collection		
Wavelength (Å)	1.1	1.54
Resolution (Å)	50.00-1.41 (1.43-1.41)	50.00-1.80 (1.89-1.80)
Unique reflections	27015 (1188)	13250 (1805)
Completeness	0.983 (0.872)	1.000 (1.000)
Redundancy	6.1 (4.4)	12.6 (5.0)
Average $I/\sigma(I)$	14.2 (5.4)	33.6 (6.1)
R_{merge}	0.050 (0.258)	0.022 (0.156)
Wilson B ($Å^2$)	27.2	26.1
Space group	$P2_1$	$P2_1$
Cell dimensions (Å ; °)	a = 31.11 ; α = 90.00 b = 60.45 ; β = 93.81 c = 38.31 ; γ = 90.00	a = 31.11 ; α = 90.00 b = 60.38 ; β = 94.24 c = 38.47 ; γ = 90.00
Refinement		
Resolution (Å)	38.22-1.40 (1.44-1.40)	25.05-1.80 (1.80-1.85)
No. of non-hydrogen atoms	1252	1181
No. of unique reflections	27001 (1560)	13205 (989)
No. of reflections in test set	1356 (73)	1306 (96)
R_{work}	0.1654(0.30)	0.161 (0.18)
R_{free}	0.188 (0.32)	0.220 (0.25)
RMS from ideal geometry		
Bonds (Å)	0.019	0.016
RMS angles (°)	1.94	1.54
Average B-factors ($Å^2$)		
Protein	23.6	20.8
Solvent	31.4	26.0
Ion	20.3	23.4
Ramachandran plot		
Most favored (%)	100 (87.7)	100 (87.7)
Additionally allowed (%)	13 (11.4)	13 (11.4)
Disallowed (%)	1 (0.9)	1 (0.9)
No. of residues excluding Gly, Pro and termini	114	114
Total no. of residues	129	129
PDB accession code	4DFA	3V2T
RMSD (Å) from Δ+PHS	0.28	0.31
Main chain only	0.12	0.14

Protein	Δ+PHS I92A/V74A	Δ+PHS I92A/L103A
Crystallization conditions		
Buffer	25mM K phosphate	25mM K phosphate
pH	6	9
Temperature (K)	277	277
Precipitant	20% (w/v) MPD	20% (w/v) MPD
Additives	pdTp, $CaCl_2$	pdTp, $CaCl_2$
Data collection		
Wavelength (Å)	1.1	1.1
Resolution (Å)	50.00-1.63 (1.66-1.63)	50.00-1.40 (1.42-1.40)
Unique reflections	17867 (882)	27568 (1382)
Completeness	0.997 (0.974)	0.990 (1.000)
Redundancy	12.1 (9.8)	6.3 (6.2)
Average $I/\sigma(I)$	16.8 (11.3)	13.4 (9.1)
R_{merge}	0.051 (0.285)	0.049 (0.219)
Wilson B ($Å^2$)	33.2	24.2
Space group	$P2_1$	$P2_1$
Cell dimensions (Å ; °)	a = 31.05 ; α = 90.00 b = 60.52 ; β = 93.85 c = 38.60 ; γ = 90.00	a = 31.14 ; α = 90.00 b = 60.36 ; β = 93.12 c = 38.36 ; γ = 90.00
Refinement		
Resolution (Å)	38.51-1.63 (1.67-1.63)	31.1-1.40 (1.44-1.40)
No. of non-hydrogen atoms	1175	1252
No. of unique reflections	17849 (1194)	27544 (1933)
No. of reflections in test set	914 (58)	2719 (184)
R_{work}	0.177 (0.24)	0.169 (0.27)
R_{free}	0.208 (0.24)	0.199 (0.27)
RMS from ideal geometry		
Bonds (Å)	0.015	0.014
RMS angles (°)	1.87	1.68
Average B-factors ($Å^2$)		
Protein	34.7	20.0
Solvent	36.5	28.1
Ion	28.4	20.8
Ramachandran plot		
Most favored (%)	101 (88.6)	100 (87.7)
Additionally allowed (%)	12 (10.5)	13 (11.4)
Disallowed (%)	1 (0.9)	1 (0.9)
No. of residues excluding Gly, Pro and termini	114	114
Total no. of residues	129	129
PDB accession code	4DU9	3SXH
RMSD (Å) from Δ+PHS	0.27	0.34
Main chain only	0.13	0.15

Protein	Δ+PHS I92A/L125A	Δ+PHS L25A/L36A/I92A
Crystallization conditions		
Buffer	25mM K phosphate	25mM K phosphate
pH	7	9
Temperature (K)	277	277
Precipitant	20% (w/v) MPD	20% (w/v) MPD
Additives	pdTp, $CaCl_2$	pdTp, $CaCl_2$
Data collection		
Wavelength (Å)	1.1	1.1
Resolution (Å)	50.00-1.47 (1.50-1.47)	50.00-1.65 (1.68-1.65)
Unique reflections	23894 (1027)	17188 (836)
Completeness	0.985 (0.862)	0.999 (0.994)
Redundancy	5.9 (3.4)	6.4 (5.8)
Average $I/\sigma(I)$	19.3 (8.3)	16.6 (13.0)
R_{merge}	0.054 (0.169)	0.042 (0.135)
Wilson B ($Å^2$)	29.5	28.5
Space group	$P2_1$	$P2_1$
Cell dimensions (Å ; °)	a = 31.07 ; α = 90.00 b = 60.40 ; β = 93.50 c = 38.48 ; γ = 90.00	a = 31.16 ; α = 90.00 b = 60.00 ; β = 95.06 c = 38.88 ; γ = 90.00
Refinement		
Resolution (Å)	32.41-1.47 (1.51-1.47)	38.73-1.65 (1.69-1.65)
No. of non-hydrogen atoms	1225	1217
No. of unique reflections	23838 (1449)	17170 (1203)
No. of reflections in test set	2397 (162)	869 (55)
R_{work}	0.174 (0.24)	0.179 (0.26)
R_{free}	0.213(0.27)	0.218 (0.28)
RMS from ideal geometry		
Bonds (Å)	0.018	0.020
RMS angles (°)	1.84	1.98
Average B-factors ($Å^2$)		
Protein	24.0	24.5
Solvent	31.3	31.2
Ion	22.4	23.6
Ramachandran plot		
Most favored (%)	98 (86.0)	100 (87.7)
Additionally allowed (%)	15 (13.2)	13 (11.4)
Disallowed (%)	1 (0.9)	1 (0.9)
No. of residues excluding Gly, Pro and termini	114	114
Total no. of residues	129	129
PDB accession code	4DGZ	4K2K
RMSD (Å) from Δ+PHS	0.35	0.38
Main chain only	0.14	0.20

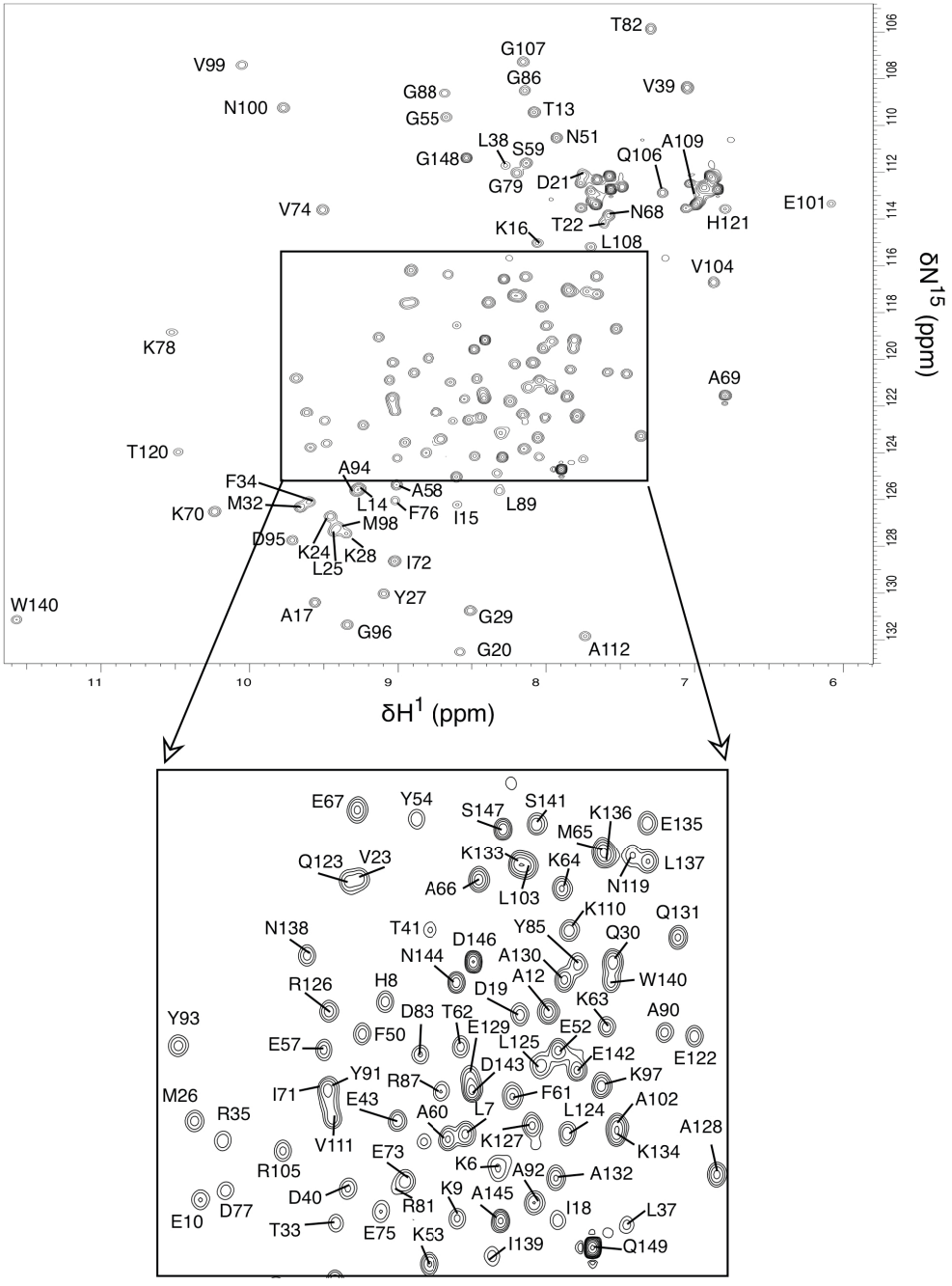
Protein	Δ+PHS L25A/V66A/I92A
Crystallization conditions	
Buffer	25mM K phosphate
pH	6
Temperature (K)	277
Precipitant	20% (w/v) MPD
Additives	pdTp, $CaCl_2$
Data collection	
Wavelength (Å)	0.9790
Resolution (Å)	50.00-1.75 (1.78-1.75)
Unique reflections	14281 (712)
Completeness	0.996 (1.000)
Redundancy	6.3 (6.5)
Average *I/σ(I)*	21 (17.7)
R_{merge}	0.043 (0.114)
Wilson B ($Å^2$)	34.3
Space group	$P2_1$
Cell dimensions (Å ; °)	a = 31.08 ; α = 90.00 b = 60.04 ; β = 94.88 c = 38.48 ; γ = 90.00
Refinement	
Resolution (Å)	38.34-1.75 (1.79-1.75)
No. of non-hydrogen atoms	1211
No. of unique reflections	14265 (982)
No. of reflections in test set	717 (44)
R_{work}	0.181 (0.19)
R_{free}	0.219 (0.19)
RMS from ideal geometry	
Bonds (Å)	0.019
RMS angles (°)	1.95
Average B-factors ($Å^2$)	
Protein	28.5
Solvent	34.3
Ion	23.7
Ramachandran plot	
Most favored (%)	101 (88.6)
Additionally allowed (%)	12 (10.5)
Disallowed (%)	1 (0.9)
No. of residues excluding Gly, Pro and termini	114
Total no. of residues	129
PDB accession code	4KD3
RMSD (Å) from Δ+PHS Main chain only	0.34 0.20

Suppl. Figure 1: NMR assignments of (A) V66A/I92A, (B) I92A/L125A and (C) L25A/V66A/I92A, acquired as described in the methods section.

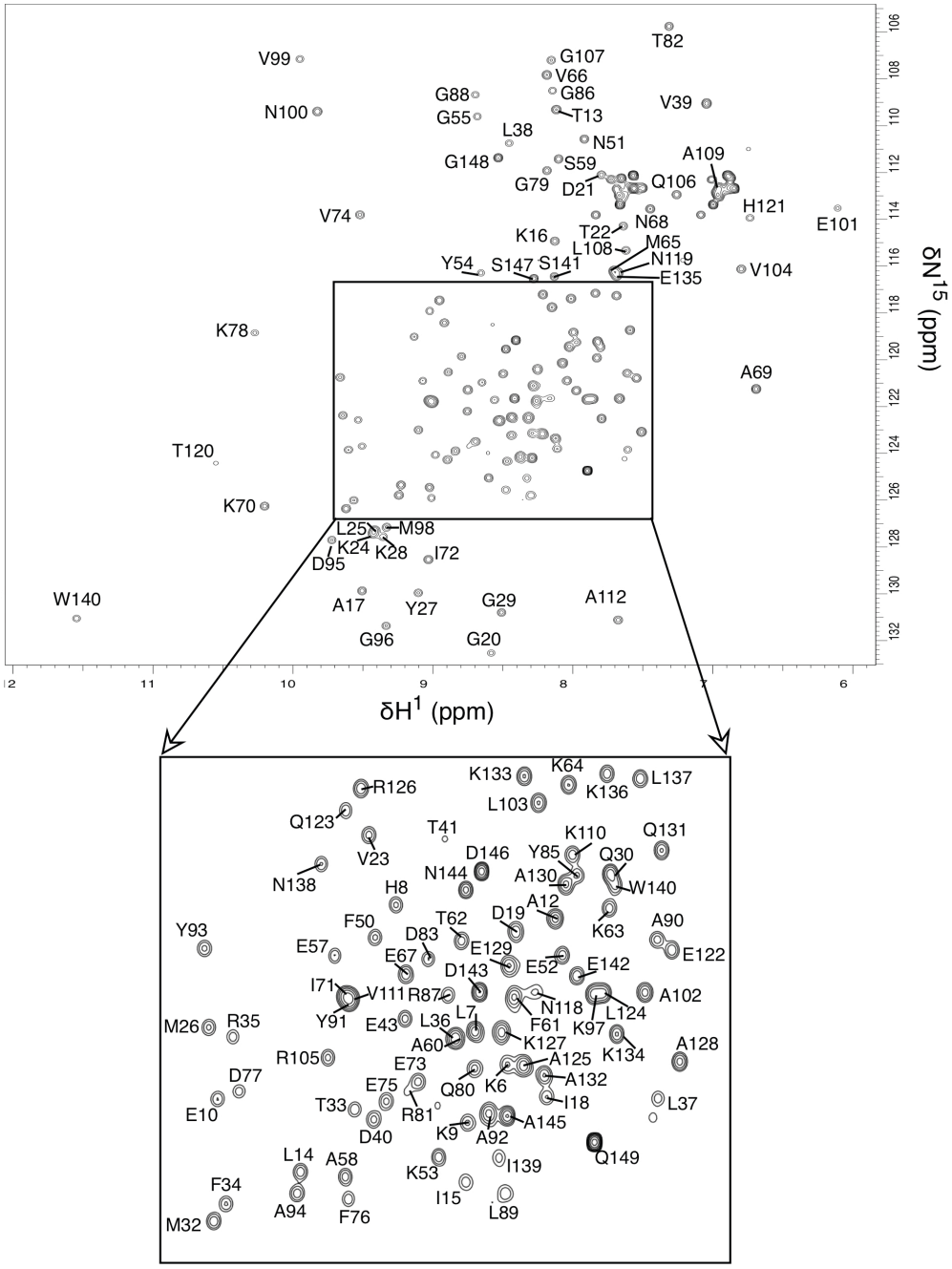
Suppl. Figure 2: Comparison between (A) $\Delta G°$ and (B) ΔV measured by fluorescence for variants with two and three Ala substitutions and the value expected by linearly adding the values measured for variants with a single Ala substitution.

Suppl. Figure 3: Changes in chemical shifts with respect to the parent protein at 1 bar in (A) the ^{15}N dimension and (B) the ^{1}H dimension. V66A/I92A is shown in red, I92A/L125A in green and L25A/V66A/I92A in blue. (C) Changes in chemical shifts between 1bar and 500 bar, between variants and the parent protein. Structural mapping of amides with the largest differences with Δ+PHS in the ^{15}N dimension at 1bar are shown in (D) red for variant V66A/I92A, (E) green for I92A/L125A, and (F) blue for L25A/V66A/I92A. Yellow spheres highlight the substitution sites.

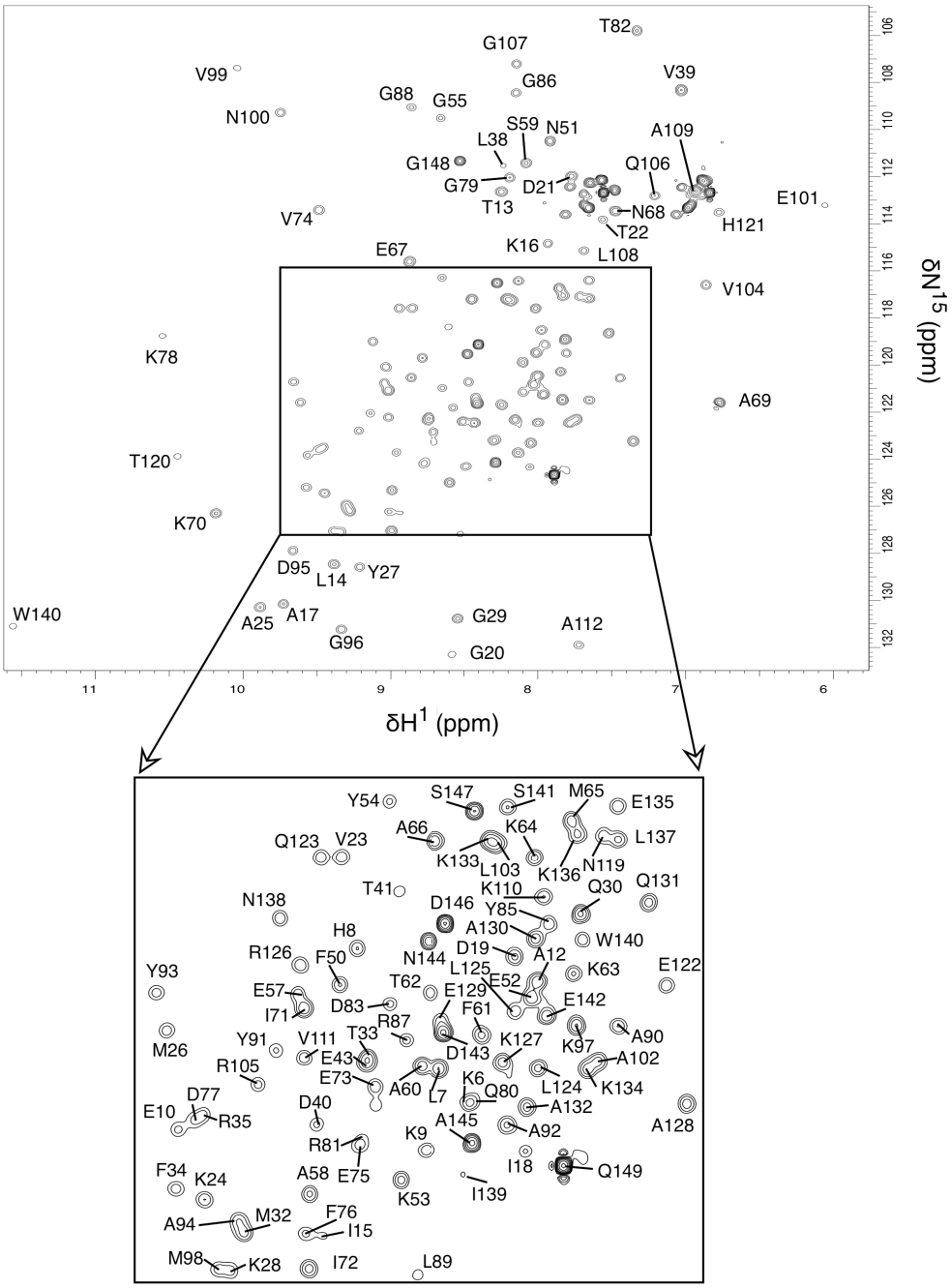
Suppl. Figure 4: Correlation plot of ΔV and calculated cavity volume using McVol and (A) a 1.4 Å probe, (B) a 1.2 Å probe. In red is Δ+PHS, in open circles are the previously published variants with one substitution to Ala, and in filled circles are the variants with two or three substitutions to Ala. In orange are the mean cavity volumes obtained for variants L36A/I92A and L25A/L36A/I92A from 100 representative frames of 10 ns long MD simulations. The dotted line represents a line with unity slope. Linear fits of data points, excluding values obtained with MD, yield intercept and slope values of (A) 63 mL/mol and 0.83 and (B) 37 mL/mol and 0.80.



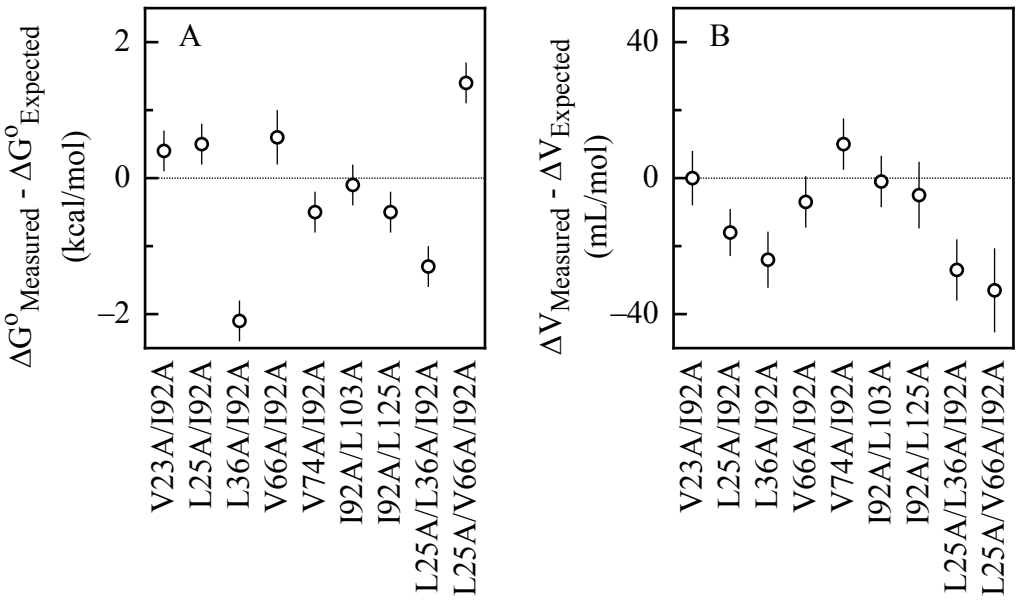
Supplemental figure 1A



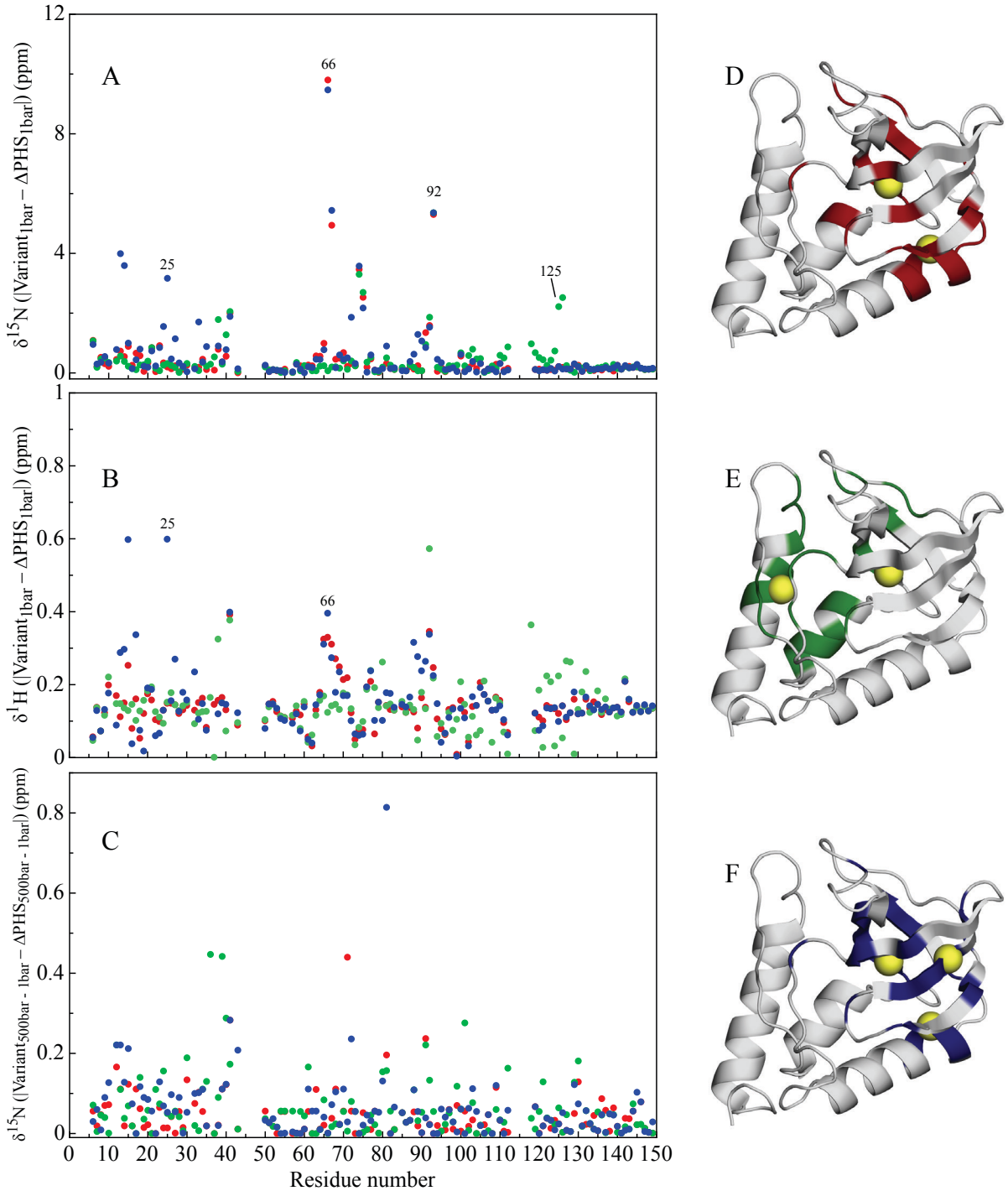
Supplemental figure 1B



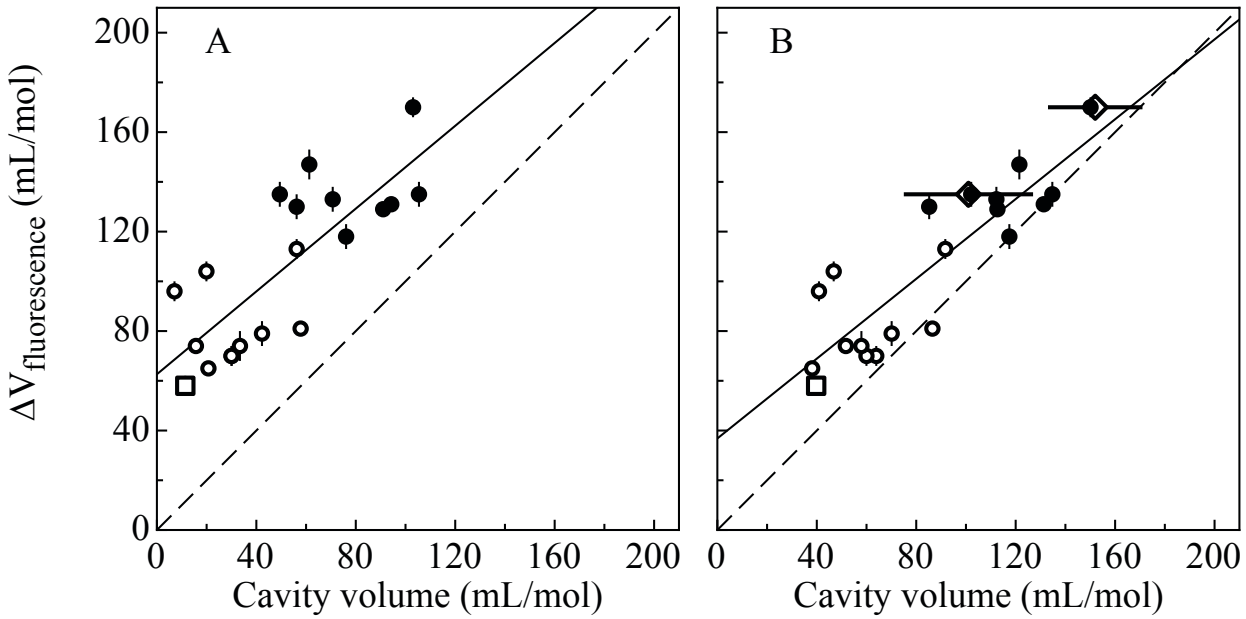
Supplemental figure 1C



Supplemental figure 2



Supplemental figure 3



Supplemental figure 4

CHAPTER 3:

FILLING CAVITIES TO DECREASE THE PRESSURE SENSITIVITY OF PROTEINS

Abstract

The pressure unfolding of proteins reflects the difference in the total volume of the folded and unfolded states (ΔV). A recent study of 20 variants of staphylococcal nuclease (SNase) with artificial cavities engineered in the folded state by substitution of core positions with Ala found a linear correlation between ΔV measured experimentally and the volume of internal cavities observed in crystal structures in proteins. The relationship demonstrates that the volume of internal cavities makes a substantial contribution towards the pressure unfolding of SNase and that contributions from other factors, such as electrostriction, are small by comparison. However, extrapolation suggests that a variant of SNase in which the internal cavities were filled would still have $\Delta V > 0$. This was examined here using wild type SNase, which has a small, naturally occurring cavity in its hydrophobic core. Six variants were engineered to attempt to fill the cavity observed in its crystal structure by substituting core residues lining the cavity with residues that increased the volume of the side chain at a site. The crystal structures of the variants confirmed the decrease in cavity volume; two of them showed no detectable cavity when inspected using a probe the size of a water molecule. Some structures show clear backbone changes localized to the region near the substitutions, highlighting the subtle structural distortions that can be induced by bulky buried side chains. The variants with localized structural distortions report ΔV values comparable to those of the parent protein, indicating that these substitutions generate more packing defects than they eliminate. Pressure unfolding experiments report a decrease in ΔV of about 30 mL/mol for the quadruple variant V23I/V66I/V74I/V99I, almost a 40% decrease in volume relative to the parent protein. The residual ΔV of approximately 60 mL/mol likely

represents the volume of cavities that are a consequence of the fluctuations inherent to this protein, or a contribution from electrostriction, or both.

Introduction

The interior of most proteins is densely packed. However, internal cavities are not uncommon and they are often essential for function, acting as determinants of the conformational fluctuations that govern allosteric signaling (1–3), ligand binding (4), and folding pathways (5). In 1914, Bridgman unexpectedly discovered that proteins are unfolded by high hydrostatic pressure (6). By Le Châtelier's principle, this implied that a difference in volume (ΔV) exists between the native and unfolded states. The physical origins of ΔV are not entirely clear (7–14). The volume of cavities present in the folded state that are eliminated upon unfolding is thought to make the dominant contribution to ΔV. This was demonstrated with a recent systematic study of variants of staphylococcal nuclease (SNase) variants with artificial internal cavities engineered by substitution of internal positions with Ala (9) (Caro et al., manuscript in preparation). The data from that study showed that a near-linear relationship exists between the volume of cavities observed in crystal structures and the ΔV measured experimentally. Extrapolation of that line to zero cavity volume led to a $\Delta V > 0$, indicating that even if a protein does not contain any cavities, contributions from other factors must exist that still make it sensitive to pressure. To further demonstrate the role of internal cavities as determinants of the pressure-sensitivity of proteins we have attempted to eliminate the pressure sensitivity of SNase by filling the dominant internal cavity with site-directed mutagenesis.

Previous studies evaluated the contribution from naturally occurring cavities found in the hydrophobic interior of proteins by using substitutions designed to fill the cavities (15–18). SNase has a small, naturally occurring cavity in its hydrophobic interior, large enough to hold a single water molecule. The V66L variant of SNase

attempted to fill the cavity by extending one of the side chains lining the cavity; however, the ΔV of this variant was larger than that of the parent protein (15). In contrast, the V103L variant of the c-Myb R2 domain showed a large shift in its pressure unfolding profile consistent with a decreased pressure sensitivity, although no ΔV was determined (16). The crystal structure of the A98L variant of T4 lysozyme showed that a Leu residue intended to fill a nearby cavity managed instead to distort an adjacent helix leading to an increase in ΔV (17, 18). The present work evaluates systematically the role of the cavity in SNase as the main determinant of its pressure unfolding. To this end, variants were engineered with substitutions from small side chains to larger side chains, intended to fill the cavity. The V23M, V23M/L36F, V23M/T62F, V23M/L25F/T62F, V23M/L36F/V66F and V23I/V66I/V74I/V99I variants contain substitutions at the internal positions that line the cavity. Crystal structures of all variants were determined to evaluate the effect of the substitutions on the overall structure and on the volume of cavities. Pressure unfolding experiments monitored by Trp fluorescence were performed to measure the effect of the substitutions on ΔV.

The results are analyzed in the context of a previous systematic study of Ala-containing variants of SNase in which an increase in the size of the main internal cavity showed a concomitant increase in ΔV (9). These previous studies were performed on the highly stable Δ+PHS variant of SNase that could withstand up to three substitutions to Ala without unfolding. However, the high stability of the background protein (ΔG° = 11.9 kcal/mol) led to the use of high molar concentration (2 M) of the chemical denaturant GdnHCl to bring the unfolding transition to the range of pressures accessible by the experimental setup (up to 3000 bar). The present study uses the true wild-type

(WT) SNase (ΔG° = 4.6 kcal/mol) and variants thereof, which require little to no addition of chemical denaturant to unfold below 3000 bar. The ΔV of WT SNase (ΔV = 79 to 92 mL/mol) (Table 2) is larger than that of Δ+PHS by 21-34 mL/mol. The difference in ΔV can be rationalized from the five surface substitutions and a truncation of 6 amino acids in a highly dynamic loop that make Δ+PHS the more stable background. Structural superposition of the two proteins shows no significant change in the volume of the main cavity, indicating that any cavity filling substitutions that reduce the ΔV of WT SNase should do the same in the Δ+PHS background.

The importance of specific packing arrangements in determining protein stability has been highlighted previously (19, 20). Optimal packing will tend to minimize cavity volumes (19) but will also determine the structural fluctuations that enable transient packing defects and transient access to the protein interior from bulk (21) as well as the entropy of internal side chains (22–24) and the residence time of waters buried in the protein interior (25–27). These processes must be sensitive to substitutions that alter the packing arrangement of the protein core and appear to be responsible for the decrease in stability observed whenever small-to-large substitutions attempt to fill internal cavities (28). In parallel to the set of SNase variants used in this study to fill cavities, another set of variants of WT SNase was engineered with isosteric substitutions in the protein interior, resulting in variants L25I/I72L, V66I/I72V, V23L/L25V/V66I/I72V, V23I/V66I/I72V/I92V and L25I/V66I/I72L/I92V. These substitutions rearranged the protein interior without changing the net van der Waals volume of the side chains nor their hydrophobic character. The isosteric variants allowed us to probe how the packing

arrangement that determines the dynamic processes of the protein interior relate to the volume of proteins and their cavities.

Results

Crystal structures. High resolution crystal structures were obtained for the V23M, V23M/L36F, V23M/T62F, V23M/L25F/T62F, V23I/V66I/V74I/V99I, V23L/L25V/V66I/I72V and V23I/V66I/I72V/I92V variants of SNase (Fig. 1 and Suppl. Table 1). The structure of V23M was the only structure for which two molecules were observed in the asymmetric unit. The two molecules were almost identical to each other, with an overall root-mean-square deviation (RMSD) value below 0.1 Å and no difference in the conformation of the internal side chains.

Superposition of the variants onto the parent protein revealed changes in side chain conformation adjacent to the substitution sites. In particular, the structure of the isosteric variant V23L/L25V/V66I/I72V showed disorder at positions 23, 36, 66 and 92, not unlike the disorder observed for the previously published structures of L25I/I72L (PDB ID 2F0N) and V66I/I72V (PDB ID 2F0O) (29, 30). For the V23M/L25F/T62F and V23M/L36F variants, calculations of RMSD highlighted a clear distortion localized around the substitution sites, along β-2 and β-3 (residues 23-30 and 33), and at residues 17, 23, 24, 30 and 33, respectively (Fig 1C and D). No water molecules nor electron density suggestive of waters were found within any of the cavities.

Cavity volume calculations. The volumes of cavities found in the crystal structures were calculated with CASTp (31) and McVol (32), which use Delaunay triangulation and Montecarlo methods, respectively, to identify and calculate cavity volumes. Visual inspection confirmed that the cavities identified by the algorithms were near the substitution sites. Because the algorithms use inherently different approaches, the

absolute values of cavity volumes will also differ. A probe of 1.4 Å in CASTp will detect similar cavities to those found by a probe of 1.2 Å in McVol. To avoid bias from different probe sizes, cavity volume calculations were run with both algorithms and with probes of both radii 1.2 Å and 1.4 Å (Table 3).

With a probe of 1.4 Å and using CASTp, the volume of the naturally occurring cavity in the core of WT SNase was 45 mL/mol. Analysis of the cavity volumes in structures of variants with small-to-large side chain substitutions with a probe the size of a water molecule (1.4 Å radius) showed either a reduction in the volume or the complete elimination of the 45 mL/mol cavity observed in the structure of WT (Fig. 1A). The range of cavity volumes calculated for the variants spaned from 33 mL/mol for V23M/T62F to 0 mL/mol for both V23M/L36F and V23M/L25F/T62F (Table 3). The V23M and V23I/V66I/V74I/V99I variants showed smaller cavities of 20 and 12 mL/mol, respectively (Fig 1E & 1F). The structure of the V23M variant showed two molecules in the asymmetric unit, and the second molecule (chain B) displayed a slightly smaller cavity of 13 mL/mol.

Variants with isosteric substitutions should, in principle, not change the internal void volume. However, variants L25I/I72L, V66I/I72V, V23L/L25V/V66I/I72V and V23I/V66I/I72V/I92V displayed cavities of volumes that range from 54 to 22 mL/mol.

Thermodynamic stability. Thermodynamic stability ($\Delta G°$) was measured by monitoring the fluorescence of Trp-140 as a function of guanidine hydrochloride (GdnHCl) concentration. None of the 9 variants studied showed an increase in stability with respect to the parent protein and most showed a decrease (Table 1 and Fig. 2). This is consistent

with previous findings in variants of T4 lysozyme where small-to-large substitutions were used to attempt to fill cavities (28) and indicates that the substituted side chains disrupt favorable interactions. The lowest stabilities of 0.6 – 1.7 kcal/mol were reported by variants V23M/L36F, V23M/T62F and V23M/L25F/T62F, which involved substitutions to the bulky Phe side chain. Stabilities as low as 1.1 kcal/mol were also reported by the isosteric variants V23L/L25V/V66I/I72V and V23I/V66I/I72V/I92V, consistent with the stability changes measured for the same variants by Chen et al. (19). These large energetic penalties should not reflect increases in cavity volume, but must instead report on loss of favorable interactions with protein surface as well as an increased entropic penalty upon folding due to changes in the dynamics of internal side chains and of local fluctuations of the protein backbone (33).

Variants with substitutions to larger side chains report *m*-values that were always lower than WT SNase by up to 1.9 kcal/mol*M. This contrasts strongly with the increase by up to 1.9 kcal/mol*M observed for the previously studied variants with 1, 2 and 3 Ala substitutions (9) (Caro et al., manuscript in preparation). The isosteric variants retained *m*-values within 0.3 kcal/mol*M of that of WT, consistent with a constant amount of solvent-accessible surface exposed upon unfolding.

Pressure unfolding monitored by Trp fluorescence. Pressure unfolding measurements were generally performed in the presence of a small amount of GdnHCl to lower the stability of the protein and bring the pressure-unfolding transition down to pressures accessible with the instrumentation available. ΔV for the WT protein, the difference in the volume of folded and unfolded states ($\Delta V = V_{unfolded} - V_{folded}$) was 79-92 mL/mol,

depending on the conditions of the experiment. It is likely that this range in ΔV values reflects a fitting artifact owing to the limited folded and unfolded baselines in the raw data. Previous experiments indicate that GdnHCl does not have any measurable effect on the ΔV of SNase, but similar denaturants can influence the pressure behavior of proteins (9, 34–37). In this work, using pH to destabilize the protein instead of GdnHCl returned the same ΔV for WT SNase (Fig 2A and B).

The V23I/V66I/V74I/V99I variants reported a ΔV value of 59-63 mL/mol, which is lower than the ΔV of the WT by 16-33 mL/mol (Table 2 and Fig 2C). The V23M, V23M/L36F, V23M/T62F and V23M/L25F/T62F variants on the other hand report ΔV values of 78-82, 89, 83-84 and 86 mL/mol, respectively, which are comparable to the ΔV of 79-92 mL/mol of the WT (Table 2 and Fig 2F).

Isosteric variants were pressure unfolded and compared to WT. Isosteric substitutions do not change the net volume of atoms in the protein core and are thus expected to report identical ΔV values. However, these variants have ΔG°'s that differ from WT by up to 3.5 kcal/mol. The variants L25I/I72L, V66I/I72V, V23I/V66I/I72V/I92V, V23L/L25V/V66I/I72V and L25I/V66I/I72L/I92V reported ΔV values of 87-92, 79-84, 81-89, 94 and 84 mL/mol, respectively, all of which were within error of the 79-92 mL/mol measured for WT (Fig 2C and D).

Discussion

Pressure sensitivity of SNase is reduced by filling cavities. The pressure sensitivity of proteins has been shown to correlate well with the volume of internal cavities in proteins (9, 18) (Caro et al., manuscript in preparation). In particular, the near-linear relationship found between the ΔV of 20 variants of SNase with increasingly large cavities and the cavities found in the crystal structures highlights the dominant role of cavities in determining the pressure sensitivity of proteins (Caro et al., manuscript in preparation). However, this relationship predicts that a protein without any cavities should still report a $\Delta V > 0$, suggesting that contributions to ΔV from other physical phenomena might be significant, such as electrostriction due to differential solvation between the folded and unfolded states (11, 12, 38), the effect of pressure on the structure of bulk water (7), or the contribution from the second-order effects electrostriction (39) and compressibility (40). This question has been previously addressed by using small-to-large substitutions that attempt to fill internal cavities (15–18). However, in none of these works was this approach successful in reducing the ΔV of the protein as measured by pressure unfolding (17, 18). The studies were simply not systematic enough to evaluate the effect of small-to-large substitutions on the protein structure, and were limited to the sometimes incomplete pressure unfolding of only one protein (15, 16). These issues are overcome in the present work by evaluating systematically the effect of different types of substitutions, including the side chains Ile, Leu, Met and Phe, and of substitutions at various positions throughout the protein, resulting in a total of 9 variants of SNase. This approach led to variant V23I/V66I/V74I/V99I, which reported the lowest ΔV of the

entire set of 9 variants of WT SNase (Table 2). Its ΔV of 59 to 63 mL/mol represents about a 30% decrease from the 79 to 92 mL/mol measured for WT.

The Val-to-Ile substitution is arguably the least disruptive of the substitutions examined. It extends the side chain by a single methylene unit and unlike a Val-to-Leu substitution (15, 16), it preserves the original packing arrangement of the β-branched side chain and adds a relatively unrestricted bond. Four such substitutions at four different sites have a much greater ability to avoid steric clashes than, for example, a Val-to-Phe substitution, which introduces an equal number of carbon atoms but with very different configuration and relaxational properties. The structure of the V23I/V66I/V74I/V99I variant showed no detectable structural distortions and a very low overall RMSD of 0.38 relative to the structure of the WT protein (PDB ID 1SNC) (Table 2), consistent with negligible changes in the overall structure. The substitutions V23I, V66I and V99I project the δ-carbon of Ile toward the cavity and reduce its volume significantly. To do so, Ile-66 adopts a rotameric state that does not coincide with that of Val-66. Substitution V74I projects the δ-carbon of Ile away from the protein interior and instead packs against two Phe (Phe-34 and Phe-76) and one Tyr (Tyr-27) side chains that do not pack together efficiently.

The V23M, V23M/L36F, V23M/T62F and V23M/L25F/T62F variants, on the other hand, all reported ΔV values within error of that of WT SNase, despite showing reduced or even no detectable cavities in the crystal structures. Close inspection of their structures revealed that the regions with the substitution sites also show high RMS deviations as compared to the WT structure. This is similar to the findings of variants of T4 lysozyme with small-to-large substitutions, which led to structural deviations from the

parent protein (18, 28, 41–43). The observed structural distortions appear to release the strain introduced by the substitutions in order to accommodate the bulky side chains that were introduced in the protein core. When a smaller probe of 1.2 Å was used to inspect the structures, smaller cavities appear adjacent to the substitution sites (Table 3), indicating that the cavities have only been partially filled and partitioned into smaller cavities. The structural distortions observed for two of the variants V23M/L36F and V23M/L25F/T62F move strands β-2 and β-3 by more than 0.5 Å from their location in the WT structure, introducing larger interstitial volumes between each atom that composes the interface with β-1 strand. The near-linear relationship reported previously (Caro et al., manuscript in preparation) between measured ΔV and cavity volumes measured from crystal structures suggests that such an increase in interstitial volumes generated by structural distortions will contribute significantly to ΔV and is consistent with an increase in ΔV observed in other proteins with small-to-large substitutions (17, 18). This conjecture implies that the pressure sensitivity of proteins must scale with protein size because the interstitial volume will be proportional to the number of buried atoms, which will be proportional to the size of the protein. This appears to be at odds with the conclusions from a study of the pressure unfolding of the ankyrin domain of the Notch receptor (Nank) and variants with deleted repeats (8). Close inspection of the data reveals, however, that the constructs involving repeats 2 through 7, 3 through 7, and 4 through 7 reported ΔV values of 55, 51 and 39 mL/mol, respectively, consistent with a decrease proportional to protein size. The changes measured in ΔV for variants of Nank are not sufficiently large to corroborate this assumption.

Over-packing protein cores can disrupt the dynamics inherent to the protein (21–27) and lead to side chain disorder (44, 45). Increased structural fluctuations can lead to transient cavities and transient access to cavities from bulk (21) and therefore an increase in ΔV of the protein. The high ΔV values measured for the variants studied here with small-to-large substitutions in the core are consistent with an increase in structural fluctuations, particularly in the region of the substitutions (Fig. 1C and 1E).

ΔV is not sensitive to specific packing arrangements. The seclusion and packing of hydrophobic residues away from solvent has been proposed a while ago as the main drive in protein folding (46). This is evident in studies of proteins that become highly destabilized upon generation of a cavity (9, 47, 48). Perhaps less apparent is the role of specific arrangements of the protein side chains in determining protein stability (19). The preference of certain packing arrangements over others is thought to be determined by the structural fluctuations and side chain dynamics inherent to the protein (19–24, 33). Because these processes involve volumetric fluctuations that result in the transient formation of packing defects and accompanying changes in internal hydration (21, 25–27) they should also influence the response of proteins to pressure. However, the 5 isosteric variants studied here indicate that as long as the van der Waals volume of internal side chains remains constant, the ΔV of the protein will remain unchanged. The 5 isosteric variants studied here, namely L25I/I72L, V66I/I72V, V23L/L25V/V66I/I72V, V23I/V66I/I72V/I92V and L25I/V66I/I72L/I92V, reported decreases in stability relative to WT SNase but no detectable change in ΔV. Even variants V23L/L25V/V66I/I72V and L25I/V66I/I72L/I92V, which showed a decrease in stability of up to 3.5 kcal/mol,

reported a ΔV value similar to that of WT. However, the crystal structures obtained showed a clear rearrangement of the protein core that resulted in a range of cavity volumes between 22 and 54 mL/mol, and in several changes in the rotameric state as well as population of multiple rotameric states of internal hydrophobic residues (Table 3).

While the substitutions are likely to affect the coupled motions between alternative side chain conformations that are determined by the specific packing of protein cores (49–52), they might not have a significant effect on the protein backbone, as suggested by the crystal structures. The fluctuations of internal hydrophobic side chains might be are too fast and not concerted enough to induce a significant motion of larger portions of the structure of the protein (51).

Interstitial volumes constitute a baseline ΔV. A nearly linear relationship was established between the experimental ΔV and the volume of cavities that yielded an intercept of 45 mL/mol and a slope of 0.67. The relationship predicts that eliminating the cavity in SNase will not eliminate the protein's pressure sensitivity but will instead yield a protein with a baseline ΔV of about 50 mL/mol. This is probably due to small packing defects throughout the protein structure, since the packing efficiency of proteins is limited by the geometric constraints imposed by side chains and by the structure of the protein backbone. In the unfolded state a near-spherical molecule like water, with none of these constraints, will always be able to improve the geometric arrangement and decrease the amount of interstitial volumes. Furthermore, in the timescale of equilibrium thermodynamic experiments used to measure pressure unfolding, the highly dynamic unfolded protein can assume a very large number of conformations that, on average,

minimize any defects in how it packs in the bulk water matrix. Still, packing efficiency in the folded state can be improved at the expense of other stabilizing interactions such as optimal H-bonding angles. Indeed, the V23I/V66I/V74I/V99I variant of SNase reports a substantial decrease in ΔV that is accompanied by a decrease in thermodynamic stability.

The correlation plot defined by the previously collected set of 20 variants of SNase with increasingly large cavities can be extended by including the set of WT SNase and the 5 proteins with cavity filling substitutions studied here (Caro et al., manuscript in preparation). The 6 data points appear to fall more or less on the line defined by the 20 previously collected data points (Fig. 4). With exception of the two outliers with structural distortions V23M/L36F and V23M/L25F/T62F (data points closest to the y-axis), the four data points appear to define a line that is shifted up relative to the line defined by the 20 variants with larger cavities.

These 20 variants previously studied were engineered onto a highly stable variant of SNase called Δ+PHS, which can withstand the large energetic penalties from substitution of internal positions with Ala. This Δ+PHS variant of SNase differs from the WT form by 5 point mutations (G50F, V51N, P117G, H124L and S128A) and the truncation of the ω-loop (residues 44-49). This translates into an increase in thermodynamic stability of 7.2 kcal/mol (298 K and 0.1 M salt) relative to the WT (53). The substitutions and the truncations used to make the Δ+PHS variant are all on the surface and do not affect the overall structure of the protein; the backbone structure of the Δ+PHS SNase and of the WT protein are superimposable. The side chains that compose the hydrophobic interior also show a highly similar arrangement, with only minor differences in the rotameric state of Val-23, Leu-36 and Val-74 that are not uncommon to

observe even between two crystals of the same variant. These small changes are observed consistently across many structures of the WT and Δ+PHS proteins (9, 54, 55) and even for the negligible structural changes observed in the two chains of the WT V23M variant that give rise to a difference in cavity volume of 8 mL/mol (Table 3). They account for the slightly larger cavity of 45 mL/mol in WT than the volume of 39 mL/mol for the cavity in the Δ+PHS protein (Table 3). The deletion of residues 44-49 in the Δ+PHS protein is expected to alter the amount of interstitial volumes and to make the Δ+PHS protein less pressure sensitive than the WT protein. Indeed, Δ+PHS reports an experimental ΔV that is 21-34 mL/mol lower than WT. It is likely that this difference is strictly due to the point mutations and the truncation of 6 residues as well as to the different dynamic properties of the two proteins, while the contribution of the cavity to ΔV remains relatively the same between the two backgrounds.

To enable a more direct comparison in the ΔV values of variants in the WT and Δ+PHS backgrounds, 34 mL/mol were subtracted from the variants engineered in the WT protein. When this correction is applied and the cavity filling variants are included in the plot of measured ΔV versus cavity volumes calculated from crystal structures, the data from the variants with the filled cavity extend the linear relationship. The V23I/V66I/V74I/V99I variant, reports a ΔV of 29 mL/mol in the pressure unfolding experiment monitored with Trp fluorescence and is thus the point closest to the origin (Fig. 4).

The subtraction of 34 mL/mol from the ΔV values collected for the 6 variants engineered onto the WT SNase is a simplistic assumption that allows for comparison with the previously collected dataset of 20 proteins with increasingly large cavities

engineered onto the Δ+PHS background of SNase. However, the compiled data points for the 26 variants define a linear relationship between the calculated volumes of internal cavities and the ΔV measured by pressure unfolding (Fig. 4C and 4D). This relationship suggests that cavity volumes could account for nearly the entire ΔV of SNase and that the contribution from factors such as electrostriction must be small or even negligible.

The unexpectedly high ΔV values of V23M/L36F and V23M/L25F/T62F were interpreted in terms of defects or interstitial volume that cannot be accounted for with a probe of size comparable to that of a water molecule. Interstitial volumes could also be invoked to explain the 29 mL/mol measured experimentally with the V23I/V66I/V74I/V99I variant. In principle this value of 29 mL/mol constitutes a baseline contribution to ΔV related to unavoidable packing defects in the folded state that are minimized in the unfolded state and which scales with the size of proteins. Despite the dominant contribution from cavities, a portion of the 29 mL/mol of variant V23I/V66I/V74I/V99I could also consist of the compounded effects of other factors such as electrostriction and compressibility, as suggested for other proteins (12, 56).

Conclusions

The systematic study presented here investigated the consequences of filling internal cavities using small-to-large substitutions. The pressure sensitivity of staphylococcal nuclease (SNase) was successfully reduced by 29 mL/mol, nearly half its original value, by engineering variant V23I/V66I/V74I/V99I with small-to-large substitutions that fill the cavity observed in the crystal structure of SNase. However, 4 other variants with small-to-large substitutions reported small or no detectable changes in the ΔV of SNase. This appeared to be due to structural distortions induced by the bulky side chains that resulted in an increase in interstitial volume. Small, interstitial volumes that are not detectable using a probe the size of a water molecule must constitute a main factor in determining the pressure sensitivity of proteins, and compose a baseline ΔV value for proteins that scales with the amount of buried surface and hence the protein size. These volumes likely account for most of the 29 mL/mol of the SNase variant V23I/V66I/V74I/V99I and imply that factors such as electrostriction, compressibility, and expansivity must not contribute significantly to the pressure sensitivity of proteins. Repacking of core residues without changing the volume of side chains does not change the ΔV of the protein, suggesting that the dynamics of internal hydrophobic side chains are too fast or not concerted enough to induce significant structural fluctuations.

Acknowledgements

This work was designed and performed by José Alfredo Caro under the supervision of his advisor Dr. Bertrand García-Moreno E. (Johns Hopkins University, Baltimore, MD). It was possible thanks to a collaboration with Dr. Catherine Ann Royer (Rensselaer Polytechnic Institute, Troy, NY), who served as co-advisor in the design and execution of the experiments, Dr. Jamie L. Schlessman (U.S. Naval Academy, Annapolis, MD) who was instrumental in the production and collection of protein crystals and in the solution of crystal structures, Dr. Mariano Dellarole and Martin Fossat (Centre de Biochimie Structurale, Montpellier, France), who participated in both collection and analysis of high pressure fluorescence data.

Bibliography

1. Wiener R, Zhang X, Wang T, Wolberger C (2012) The mechanism of OTUB1-mediated inhibition of ubiquitination. *Nature* 483(7391):618–622.

2. Sennett NC, Kadirvelraj R, Wood ZA (2011) Conformational Flexibility in the Allosteric Regulation of Human UDP-α-D-Glucose 6-Dehydrogenase. *Biochemistry* 50(44):9651–9663.

3. Kadirvelraj R, Sennett NC, Polizzi SJ, Weitzel S, Wood ZA (2011) Role of Packing Defects in the Evolution of Allostery and Induced Fit in Human UDP-Glucose Dehydrogenase. *Biochemistry* 50(25):5780–5789.

4. López CJ, Yang Z, Altenbach C, Hubbell WL (2013) Conformational selection and adaptation to ligand binding in T4 lysozyme cavity mutants. *Proc. Natl. Acad. Sci. U.S.A.* 110(46):E4306–E4315.

5. Lerch MT, Horwitz J, McCoy J, Hubbell WL (2013) Circular dichroism and site-directed spin labeling reveal structural and dynamical features of high-pressure states of myoglobin. *Proc. Natl. Acad. Sci. U.S.A.* 110(49):E4714–E4722.

6. Bridgman PW (1914) The coagulation of albumen by pressure. *J. Biol. Chem.* 19(4):511–512.

7. Grigera JR, McCarthy AN (2010) The behavior of the hydrophobic effect under pressure and protein denaturation. *Biophys. J.* 98(8):1626–1631.

8. Rouget J-B, Aksel T, Roche J, Saldana J-L, Garcia AE, Barrick D, Royer CA (2011) Size and sequence and the volume change of protein folding. *J. Am. Chem. Soc.* 133(15):6020–6027.

9. Roche J, Caro JA, Norberto DR, Barthe P, Roumestand C, Schlessman JL, Garcia AE, García-Moreno E. B, Royer CA (2012) Cavities determine the pressure unfolding of proteins. *Proc. Natl. Acad. Sci. U.S.A.* 109(18):6945–6950.

10. Chalikian TV, Macgregor Jr RB (2009) Origins of pressure-induced protein transitions. *J. Mol. Biol.* 394(5):834–842.

11. Frye KJ, Perman CS, Royer CA (1996) Testing the correlation between ΔA and ΔV of protein unfolding using m value mutants of staphylococcal nuclease. *Biochemistry* 35(31):10234–10239.

12. Urbauer JL, Ehrhardt MR, Bieber RJ, Flynn PF, Wand AJ (1996) High-Resolution Triple-Resonance NMR Spectroscopy of a Novel Calmodulin-Peptide Complex at Kilobar Pressures. *J. Am. Chem. Soc.* 118(45):11329–11330.

13. Dellarole M, Kobayashi K, Rouget J-B, Caro JA, Roche J, Islam MM, García-Moreno E. B, Kuroda Y, Royer CA (2013) Probing the Physical Determinants of Thermal Expansion of Folded Proteins. *J. Phys. Chem. B* 117(42):12742–12749.

14. Hummer G, Garde S, García AE, Paulaitis ME, Pratt LR (1998) The pressure dependence of hydrophobic interactions is consistent with the observed pressure denaturation of proteins. *Proc. Natl. Acad. Sci. U.S.A.* 95(4):1552–1555.

15. Frye KJ, Royer CA (1998) Probing the contribution of internal cavities to the volume change of protein unfolding under pressure. *Protein Sci.* 7(10):2217–2222.

16. Lassalle MW, Yamada H, Morii H, Ogata K, Sarai a, Akasaka K (2001) Filling a cavity dramatically increases pressure stability of the c-Myb R2 subdomain. *Proteins Struct. Funct. Bioinforma.* 45(1):96–101.

17. Liu R, Baase WA, Matthews BW (2000) The introduction of strain and its effects on the structure and stability of T4 lysozyme. *J. Mol. Biol.* 295(1):127–145.

18. Ando N, Barstow B, Baase WA, Fields A, Matthews BW, Gruner SM (2008) Structural and thermodynamic characterization of T4 lysozyme mutants and the contribution of internal cavities to pressure denaturation. *Biochemistry* 47(42):11097–11109.

19. Chen J, Stites WE (2001) Packing Is a Key Selection Factor in the Evolution of Protein Hydrophobic Cores. *Biochemistry* 40:15280–15289.

20. Holder JB, Bennett AF, Chen J, Spencer DS, Byrne MP, Stites WE (2001) Energetics of Side Chain Packing in Staphylococcal Nuclease Assessed by Exchange of Valines, Isoleucines, and Leucines. (31):13998–14003.

21. Persson F, Halle B (2013) Transient access to the protein interior: simulation versus NMR. *J. Am. Chem. Soc.* 135(23):8735–48.

22. Fu Y, Kasinath V, Moorman VR, Nucci NV, Hilser VJ, Wand AJ (2012) Coupled motion in proteins revealed by pressure perturbation. *J. Am. Chem. Soc.* 134(20):8543–8550.

23. Lee AL, Kinnear SA, Wand AJ (2000) Redistribution and loss of side chain entropy upon formation of a calmodulin-peptide complex. *Nat. Struct. Biol.* 7(1):72–77.

24. Lee AL, Sharp KA, Kranz JK, Song X-J, Wand AJ (2002) Temperature dependence of the internal dynamics of a calmodulin-peptide complex. *Biochemistry* 41(46):13814–13825.

25. Halle B (2004) Protein hydration dynamics in solution: a critical survey. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 359(1448):1207–23; discussion 1223–4, 1323–8.

26. Halle B (1997) NMR identification of hydrophobic cavities with low water occupancies in protein structures using small gas molecules. *Nat. Struct. Biol.*

27. Nucci NV, Fuglestad B, Athanasoula EA, Wand AJ (2014) Role of cavities and hydration in the pressure unfolding of T4 lysozyme. *Proc. Natl. Acad. Sci. U.S.A.* 111(38):13846–13851.

28. Karpusas M, Baase WA, Matsumura M, Matthews BW (1989) Hydrophobic packing in T4 lysozyme probed by cavity-filling mutants. *Proc. Natl. Acad. Sci. U.S.A.* 86(21):8237–8241.

29. Chen J, Lu Z, Sakon J, Stites WE (2004) Proteins with simplified hydrophobic cores compared to other packing mutants. *Biophys. Chem.* 110(3):239–48.

30. Chen J, Lu Z, Sakon J, Stites WE (2000) Increasing the thermostability of staphylococcal nuclease: implications for the origin of protein thermostability. *J. Mol. Biol.* 303(2):125–30.

31. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* 34(suppl 2):W116–W118.

32. Till MS, Ullmann GM (2010) McVol - a program for calculating protein volumes and identifying cavities by a Monte Carlo algorithm. *J. Mol. Model.* 16(3):419–429.

33. Igumenova TI, Frederick KK, Wand AJ (2006) Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution. *Chem. Rev.* 106(5):1672–99.

34. Herberhold H, Royer CA, Winter R (2004) Effects of chaotropic and kosmotropic cosolvents on the pressure-induced unfolding and denaturation of proteins: An FT-IR study on staphylococcal nuclease. *Biochemistry* 43(12):3336–3345.

35. Ravindra R, Winter R (2004) Pressure perturbation calorimetry: a new technique provides surprising results on the effects of co-solvents on protein solvation and unfolding behaviour. *Chemphyschem* 5(4):566–571.

36. Ravindra R, Royer CA, Winter R (2004) Pressure perturbation calorimetic studies of the solvation properties and the thermal unfolding of staphylococcal nuclease. *Phys. Chem. Chem. Phys.* 6(8):1952–1961.

37. Mitra L, Smolin N, Ravindra R, Royer CA, Winter R (2006) Pressure perturbation calorimetric studies of the solvation properties and the thermal unfolding of proteins in solution-experiments and theoretical interpretation. *Phys. Chem. Chem. Phys.* 8(11):1249–1265.

38. Brun L, Isom DG, Velu P, García-Moreno E. B, Royer CA (2006) Hydration of the Folding Transition State Ensemble of a Protein. *Biochemistry* 45(11):3473–3480.

39. Vasilchuk D, Pandharipande PP, Suladze S, Sanchez-Ruiz JM, Makhatadze GI (2014) Molecular determinants of expansivity of native globular proteins: a pressure perturbation calorimetry study. *J. Phys. Chem. B* 118(23):6117–6122.

40. Seemann H, Winter R, Royer CA (2001) Volume, expansivity and isothermal compressibility changes associated with temperature and pressure unfolding of Staphylococcal nuclease. *J. Mol. Biol.* 307(4):1091–1102.

41. Bueno M, Cremades N, Neira JL, Sancho J (2006) Filling small, empty protein cavities: structural and energetic consequences. *J. Mol. Biol.* 358(3):701–12.

42. Tanaka M, Chon H, Angkawidjaja C, Koga Y, Takano K, Kanaya S (2010) Protein core adaptability: crystal structures of the cavity-filling variants of Escherichia coli RNase HI. *Protein Pept. Lett.* 17(9):1163–9.

43. Eriksson AE, Baase WA, Matthews BW (1992) Similar hydrophobic replacements of Leu99 and Phe153 within the core of T4 lysozyme have different structural and thermodynamic consequences. *J. Mol. Biol.* 229(3):747–769.

44. Wynn R, Harkins PC, Richards FM, Fox RO (1997) Comparison of straight chain and cyclic unnatural amino acids embedded in the core of staphylococcal nuclease. *Protein Sci.* 6(8):1621–6.

45. Wynn R, Harkins PC, Richards FM, Fox RO (1996) Mobile unnatural amino acid side chains in the core of staphylococcal nuclease. *Protein Sci.* 5(6):1026–1031.

46. Kauzmann W (1959) Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* 14:1–63.

47. Eriksson AE, Baase WA, Zhang XJ, Heinz DW, Blaber M, Baldwin EP, Matthews BW (1992) Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* 255(5041):178–183.

48. Xu J, Baase WA, Baldwin E, Matthews BW (1998) The response of T4 lysozyme to large-to-small substitutions within the core and its relation to the hydrophobic effect. *Protein Sci.* 7(1):158–177.

49. Fraser JS, Clarkson MW, Degnan SC, Erion R, Kern D, Alber T (2009) Hidden alternative structures of proline isomerase essential for catalysis. *Nature* 462(7273):669–73.

50. Fraser JS, van den Bedem H, Samelson AJ, Lang PT, Holton JM, Echols N, Alber T (2011) Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci. U.S.A.* 108(39):16247–16252.

51. Sharp KA, Kasinath V, Wand AJ (2014) Banding of NMR-derived methyl order parameters: Implications for protein dynamics. *Proteins* (January):1–12.

52. Halle B (2002) Flexibility and packing in proteins. *Proc. Natl. Acad. Sci. U.S.A.* 99(3):1274–9.

53. García-Moreno E. B, Dwyer JJ, Gittis AG, Lattman EE, Spencer DS, Stites WE (1997) Experimental measurement of the effective dielectric in the hydrophobic core of a protein. *Biophys. Chem.* 64(1-3):211–224.

54. Chimenti MS, Khangulov VS, Robinson AC, Heroux A, Majumdar A, Schlessman JL, García-Moreno B (2012) Structural reorganization triggered by charging of Lys residues in the hydrophobic interior of a protein. *Structure* 20(6):1071–85.

55. Schlessman JL, Abe C, Gittis A, Karp DA, Dolan MA, García-Moreno E. B (2008) Crystallographic study of hydration of an internal cavity in engineered proteins with buried polar or ionizable groups. *Biophys. J.* 94(8):3208–16.

56. Taulier N, Chalikian TV (2002) Compressibility of protein transitions. *Biochim. Biophys. Acta* 1595(1-2):48–70.

57. Castañeda CA, Fitch CA, Majumdar A, Khangulov V, Schlessman JL, García-Moreno E. B (2009) Molecular determinants of the pKa values of Asp and Glu residues in staphylococcal nuclease. *Proteins Struct. Funct. Bioinforma.* 77(3):570–88.

58. Loll PJ, Lattman EE (1989) The Crystal Structure of the Ternary Complex of Staphylococcal Nuclease, Ca2+, and the inhibitor pdTp, refined at 1.65 A. *Proteins Struct. Funct. Bioinforma.* 5:183–201.

Table 1: Thermodynamic values from GdnHCl denaturation experiments at 298 K. [a, b]

Protein	ΔG^{o}_{H2O} (kcal/mol)	$\Delta\Delta G^{o}_{H2O}$ (kcal/mol)	*m*-value (kcal/mol*M)	C_{mid} (M)
WT	4.6 (0.1)	--	6.1 (0.1)	0.8 (0.1)
V23M	4.6 (0.1)	0.0	6.1 (0.1)	0.8 (0.1)
V23I/V66I/V74I/V99I	2.7 (0.1)	-1.9	4.2 (0.1)	0.6 (0.1)
V23M/T62F	1.7 (0.1)	-2.9	4.2 (0.1)	0.4 (0.1)
V23M/L36F	1.5 (0.1)	-3.1	5.6 (0.2)	0.3 (0.1)
V23M/L25F/T62F	0.6 (0.1)	-4.0	5.6 (0.1)	0.1 (0.1)
V66I/I72V	3.3 (0.1)	-1.3	5.9 (0.1)	0.6 (0.1)
V23I/V66I/I72V/I92V	3.1 (0.1)	-1.5	5.8 (0.1)	0.5 (0.1)
L25I/I72L	2.9 (0.1)	-1.7	6.3 (0.1)	0.5 (0.1)
V23L/L25V/V66I/I72V	1.2 (0.1)	-3.4	6.1 (0.1)	0.2 (0.1)
L25I/V66I/I72L/I92V	1.1 (0.1)	-3.5	6.2 (0.1)	0.2 (0.1)

[a]All experiments were performed in 25 mM HEPES with 100 mM KCl at 25 °C, pH 7

[b]Values in parenthesis represent the error of the fit. A single experiment was performed for each protein.

Table 2: Thermodynamic values from pressure denaturation experiments at 293 K.

Protein	Conditions (M GdnHCl)	ΔG^{o}_{1bar} (kcal/mol)	$\Delta V_{experimental}$ (mL/mol)	P_{mid} (bar)
WT	0.30	3.2 (0.1)	86 (3)	1557 (511)
	0.35	3.0 (0.1)	92 (2)	1364 (295)
	0.40	2.5 (0.1)	87 (3)	1202 (394)
	0.15 pH 5.5	2.9 (0.1)	79 (2)	1526 (355)
	0.20 pH 5.5	2.8 (0.1)	89 (2)	1336 (294)
V23M	0.30	3.7 (0.2)	78 (5)	1985 (1143)
	0.35	2.6 (0.1)	82 (3)	1327 (447)
	0.40	2.3 (0.1)	78 (4)	1234 (565)
V23M/L36F	0.00	2.4 (0.1)	89 (2)	1128 (250)
V23M/T62F	0.10	2.5 (0.1)	84 (3)	1245 (415)
	0.15	2.1 (0.1)	83 (4)	1050 (467)
V23M/L25F/T62F	0.00	1.8 (0.1)	86 (5)	876 (477)
V23I/V66I/V74I/V99I	0.30	2.6 (0.1)	59 (2)	1856 (497)
	0.40	2.1 (0.1)	63 (2)	1376 (359)
L25I/I72L	0.10	2.8 (0.1)	87 (3)	1347 (441)
	0.15	2.5 (0.1)	92 (3)	1137 (363)
V66I/I72V	0.10	3.4 (0.1)	79 (3)	1801 (616)
	0.20	2.7 (0.1)	84 (4)	1345 (593)
V23I/V66I/I72V/I92V	0.10	3.0 (0.1)	81 (2)	1550 (356)
	0.20	2.5 (0.1)	89 (3)	1175 (381)
V23L/L25V/V66I/I72V	0.00	1.8 (0.1)	94 (3)	801 (255)
L25I/V66I/I72L/I92V	0.00	1.8 (0.1)	84 (4)	897 (397)

Table 3: Volume calculations.

Protein	Change in van der Waals volume (mL/mol)	Probe radius (Å)	Calculated volume (mL/mol)		PDB ID
			McVol	CASTp	
Δ+PHS	0.00	1.2	40	84	3BDC*
		1.4	11	39	
WT	0.00	1.2	53	104	1SNC*
		1.4	21	45	
L25I/I72L	0.00	1.2	57	109	2F0N*
		1.4	16	26	
V66I/I72V	0.00	1.2	32	80	2F0O*
		1.4	7	22	
V23L/L25V/V66I/I72V	0.00	1.2	56	96	4KD4
		1.4	20	54	
V23I/V66I/I72V/I92V	0.00	1.2	28	66	4R8N
		1.4	14	28	
V23M	-11.23	1.2	27 [30]	63 [80]	4QF4
		1.4	8 [9]	21 [13]	
V23M/L36F	-18.12	1.2	16	27	4K5X
		1.4	0	0	
V23M/T62F	-36.54	1.2	33	57	4K6D
		1.4	20	33	
V23M/L25F/T62F	-43.43	1.2	11	18	4K5W
		1.4	0	0	
V23I/V66I/V74I/V99I	-43.90	1.2	9	34	4ODG
		1.4	8	12	

^The values in brackets for variant V23M correspond to chain B of the asymmetric unit.

* Structures by Castañeda et al. (57), Loll & Lattman (58), and Lu, Sakon and Stites (2005).

Figure legends

Figure 1. High-resolution crystal structures of SNase variants with substitutions in internal positions. Cavities are shown in red, substituted side chains as sticks (carbon as green, sulfur as yellow). **(A)** WT SNase (PDBID 1SNC) with a box highlighting zoomed in region, **(B)** V23M (PDBID 4QF4), **(C)** V23M/L36F (PDBID 4K5X), **(D)** V23M/T62F (PDBID 4K6D), **(E)** V23M/L25F/T62F (PDBID 4K5W), **(F)** V23I/V66I/V74I/V99I (PDBID 4ODG), **(G)** V23L/L25V/V66I/I72V (PDBID 4KD4) and **(H)** V23I/V66I/I72V/I92V (PDBID 4R8N).

Figure 2. Chemical denaturation monitored by Trp fluorescence for variants with V23M, V23M/L36F, V23M/T62F, L25I/I72L, V66I/I72L, V23M/L25F/T62F, V23I/V66I/I72V/I92V, V23L/L25V/V66I/I72V, V23I/V66I/V74I/V99I, and L25I/V66I/I72L/I92V. WT SNase is highlighted in bold.

Figure 3. Normalized pressure unfolding profiles measured by following the center of spectral mass of Trp-140 fluorescence as a function of pressure. **(A)** WT SNase in 0.35M GdnHCl and pH 7 (•), and in 0.2M GdnHCl and pH 5.5 in open squares (□). **(B)** Variant L25I/V66I/I72L/I92V in 0M GdnHCl in filled circles (•) L25I/I27L in 0.1M GdnHCl in crosses (+), V66I/I72V in 0.2M GdnHCl in open squares (□), and V23I/V66I/I72V/I92V in 0.2M GdnHCl in filled triangles (▼). For comparison, the fit of WT SNase in 0.4M GdnHCl is shown as a dotted line (⋯). **(C)** Variant V23M in 0.35M GdnHCl in filled circles (•) and V23I/V66I/V74I/V99I in 0.4M GdnHCl in open squares (□). For comparison, the fit of WT SNase in 0.35M GdnHCl is shown as a dotted line (⋯). **(D)**

WT SNase in 0.30M GdnHCl and pH 7, filled circles (•), and in 0.15M GdnHCl and pH 5.5, open squares (□). **(E)** Variant V23L/L25V/V66I/I72V in 0M GdnHCl in filled circles (•), V66I/I72V in 0.1M GdnHCl in crosses (+), V23I/V66I/I72V/I92V in 0.1M GdnHCl in open squares (□), and L25I/I72L in 0.15M GdnHCl in filled triangles (▼). For comparison, the fit of WT SNase in 0.4M GdnHCl is shown as a dotted line (···). **(F)** V23M in 0.4M GdnHCl in filled circles (•), V23M/T62F in 0.1M GdnHCl in crosses, V23M/L36F in 0M GdnHCl in open squares (□), V23M/L25F/T62F in 0M GdnHCl in filled triangles (▼). For comparison, the fit of WT SNase in 0.4M GdnHCl is shown as a dotted line (···). All measurements were performed at pH 7 unless otherwise specified.

Figure 4. Correlation plot of measured ΔV versus calculated cavity volume. **(A, B)** ΔV values obtained by high pressure fluorescence are plotted against the calculated volume of cavities observed by crystallography. In open circles (o) are the values obtained for alanine variants of Δ+PHS and the closed circle (•) is the value previously measured for Δ+PHS (Caro et al., manuscript in preparation). In open triangles (△) are the values obtained from this study. The closed triangle (▲) is the value of WT from this study. Volumes were calculated using CASTp (31) and a probe of **(A)**1.4 Å and **(B)** 1.2 Å. Fits do not include variants with cavity filling substitutions and yield an intercept and slope of **(A)** 45 mL/mol and 0.67 and **(B)** -5 mL/mol and 0.75. **(C, D)** The ΔV values for WT variants have been translocated by subtracting 34 mL/mol from their original value, the difference in ΔV between the WT and Δ+PHS backgrounds, as described in the discussion section. Volumes were calculated using CASTp (31) and a probe of **(C)** 1.4 Å

and **(D)** 1.2 Å. Fits include all points yield an intercept and slope of **(C)** 40 mL/mol and 0.72 and **(D)** 13 mL/mol and 0.63.

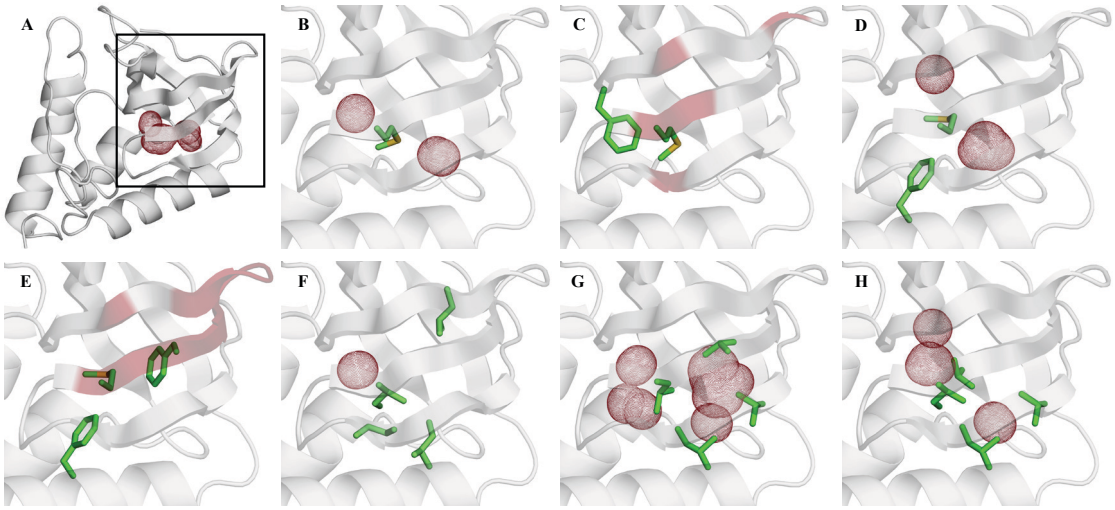


Figure 1.

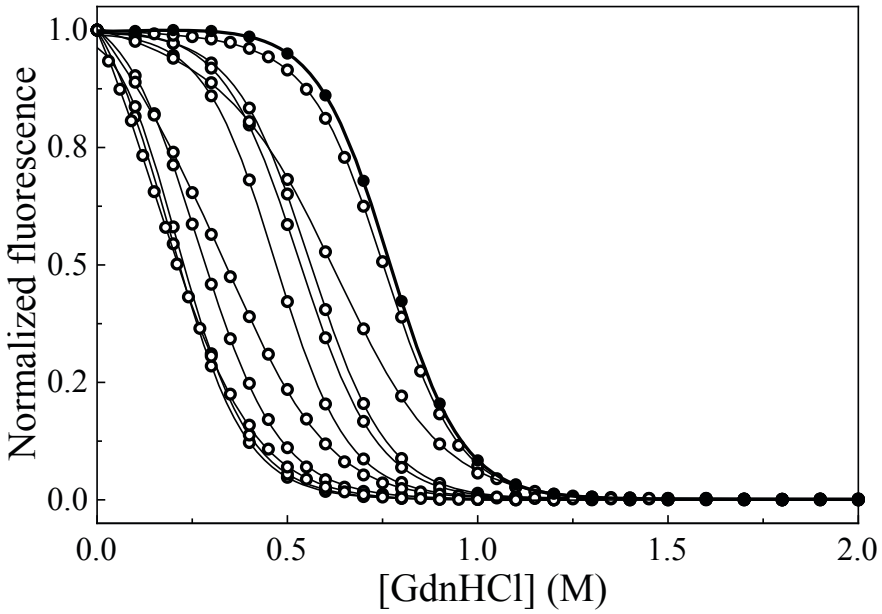


Figure 2.

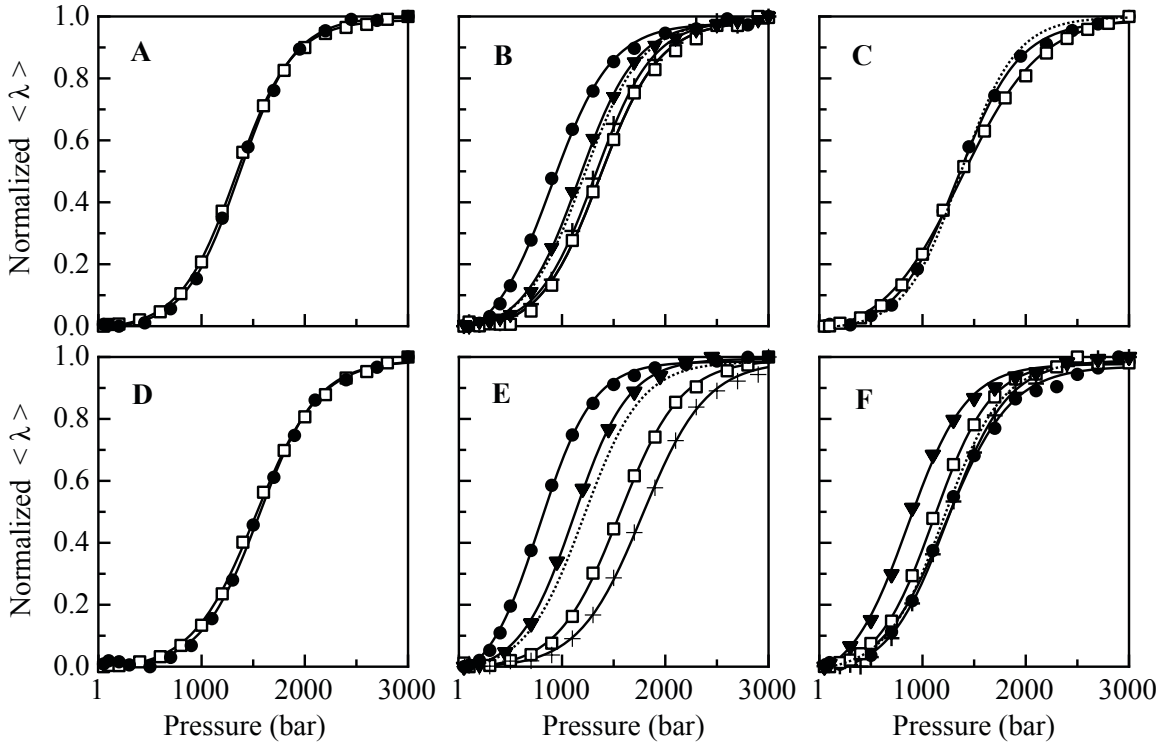

A
B
C
D
E
F
Normalized < λ >
Normalized < λ >
1.0
0.8
0.6
0.4
0.2
0.0
1
1000
2000
3000
Pressure (bar)
Pressure (bar)
Pressure (bar)

Figure 3.

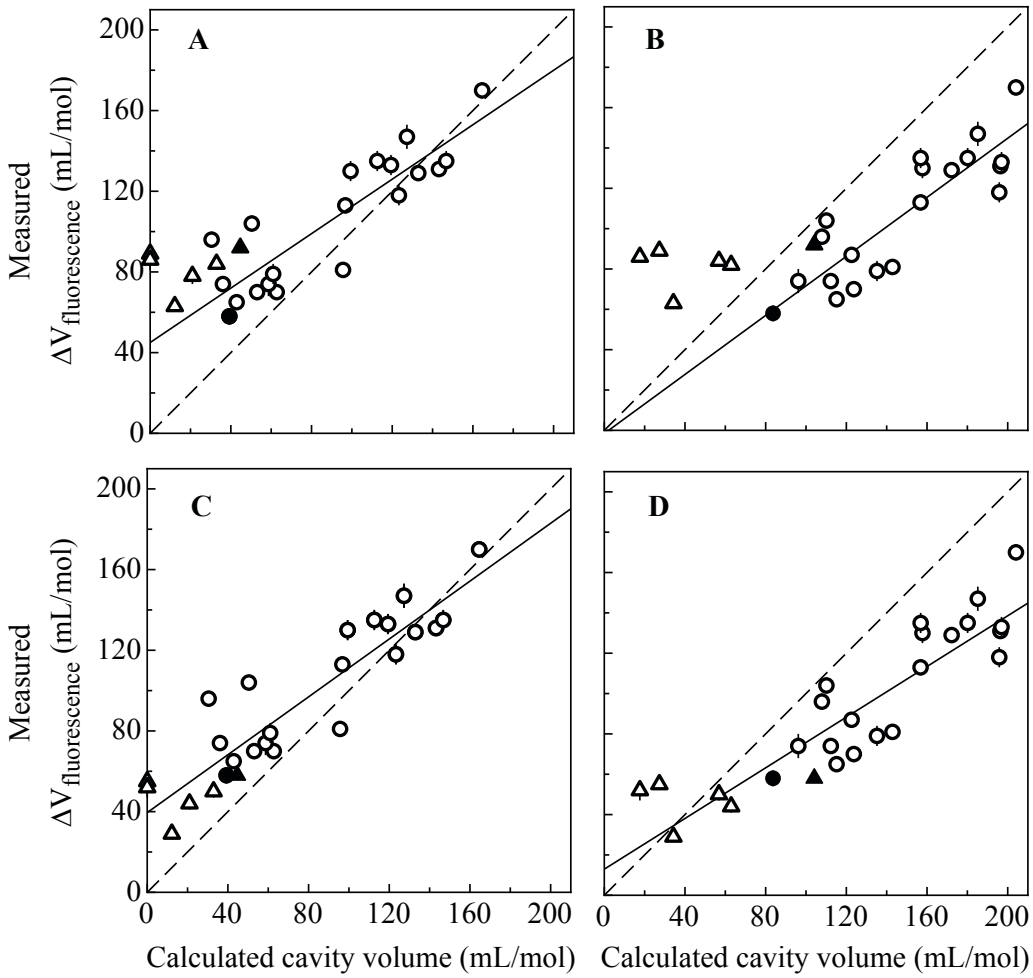

A
B
C
D
Measured
$\Delta V_{fluorescence}$ (mL/mol)
Calculated cavity volume (mL/mol)
0
40
80
120
160
200

Figure 4.

Supplemental Table 1: Crystallographic values

Crystallization conditions	V23M	V23M/L36F
Buffer	25mM potassium phosphate	25mM potassium phosphate
pH	6	9
Temperature (K)	277	277
Precipitant	32% (w/v) MPD	30% (w/v) MPD
Additives	pdTp, Calcium Chloride	pdTp, Calcium Chloride
Data collection		
Wavelength (Å)	0.979	1.1
Resolution (Å)	50.00-1.80 (1.83-1.80)[a]	50.00-1.65 (1.68-1.65)
Unique reflections	26669 (1339)	17052 (831)
Completeness	0.997 (1.000)	0.998 (0.993)
Redundancy	12.1 (12.7)	12.8 (12.5)
Average $I/\sigma(I)$	9.6 (9.0)	16.9 (9.5)
R_{merge}	0.073 (0.316)	0.038 (0.314)
Wilson B ($Å^2$)	36.6	31.7
Space group	$P4_1$	$P4_1$
Cell dimensions (Å ; °)	a = 67.53 ; α = 90.00 b = 67.53 ; β = 90.00 c = 62.82 ; γ = 90.00	a = 47.65 ; α = 90.00 b = 47.65 ; β = 90.00 c = 62.92 ; γ = 90.00
Refinement		
Resolution (Å)	38.0-1.80 (1.79-1.80)	47.65-1.65 (1.69-1.65)
No. of non-hydrogen atoms	2337	1229
No. of unique reflections	26206 (1874)	17030 (1172)
No. of reflections in test set	1329 (9 3)	863 (65)
R_{work}	0.208 (0.31)	0.187 (0.27)
R_{free}	0.240 (0.36)	0.214 (0.32)
RMSD from ideal geometry		
Bonds (Å)	0.018	0.019
RMS angles (°)	1.91	1.97
Average B-factors ($Å^2$)		
Protein	35.2	25.9
Solvent	36.1	32.7
Ion	37.1	30.8
Ramachandran plot		
Most favored (%)	194 (84.3)	100 (87.0)
Additionally allowed (%)	34 (14.7)	14 (12.2)
Disallowed (%)	2 (0.9)	1 (0.9)
No. of residues excluding Gly, Pro and termini	230	115
Total no. of residues	264	132
PDB accession code	4QF4	4K5X
RMSD (Å) from SNase	0.37	0.38
Main chain only	0.19	0.22

Supplemental Table 1 (continued)

Crystallization conditions	V23M/T62F	V23M/L25F/T62F
Buffer	25mM potassium phosphate	25mM potassium phosphate
pH	6	8
Temperature (K)	277	277
Precipitant	34% (w/v) MPD	32% (w/v) MPD
Additives	pdTp, Calcium Chloride	pdTp, Calcium Chloride
Data collection		
Wavelength (Å)	1.1	1.0
Resolution (Å)	50.00-1.55 (1.58-1.55)	50.00-1.65 (1.68-1.65)
Unique reflections	20334 (1019)	16839 (825)
Completeness	1.000 (1.000)	1.000 (1.000)
Redundancy	12.5 (12.8)	12.6 (12.9)
Average *I/σ(I)*	14.9 (8.7)	17.1 (10.8)
R_{merge}	0.065 (0.307)	0.060 (0.260)
Wilson B ($Å^2$)	31.9	34.7
Space group	$P4_1$	$P4_1$
Cell dimensions (Å ; °)	a = 47.53 ; α = 90.00 b = 47.53 ; β = 90.00 c = 62.75 ; γ = 90.00	a = 47.64 ; α = 90.00 b = 47.64 ; β = 90.00 c = 62.66 ; γ = 90.00
Refinement		
Resolution (Å)	47.53-1.65 (1.69-1.65)	47.64-1.65 (1.70-1.65)
No. of non-hydrogen atoms	1178	1243
No. of unique reflections	16830 (1198)	16811 (1192)
No. of reflections in test set	851 (73)	850 (49)
R_{work}	0.190 (0.25)	0.172 (0.21)
R_{free}	0.213 (0.30)	0.202 (0.25)
RMSD from ideal geometry		
Bonds (Å)	0.019	0.019
RMS angles (°)	1.83	1.97
Average B-factors ($Å^2$)		
Protein	23.9	30.8
Solvent	31.4	36.2
Ion	40.5	33.3
Ramachandran plot		
Most favored (%)	98 (89.9)	99 (86.1)
Additionally allowed (%)	10 (9.2)	15 (13.0)
Disallowed (%)	1 (0.9)	1 (0.9)
No. of residues excluding Gly, Pro and termini	109	115
Total no. of residues	125	132
PDB accession code	4K6D	4K5W
RMSD (Å) from SNase	0.46	0.43
Main chain only	0.29	0.24

Supplemental Table 1 (continued)

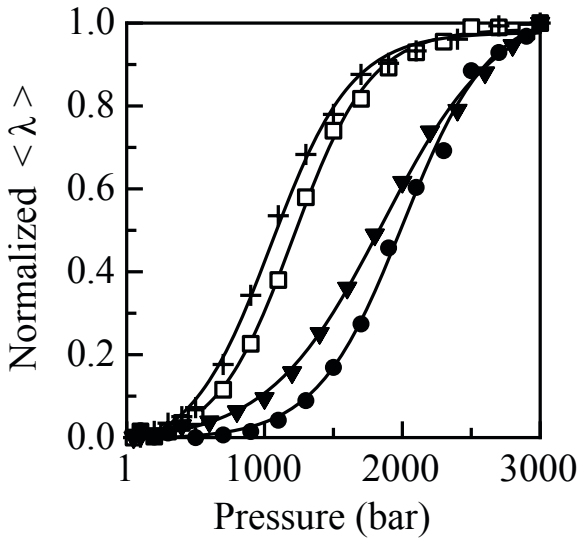
Crystallization conditions	V23I/V66I/V74I/V99I	V23L/L25V/V66I/I72V
Buffer	25mM potassium phosphate	25mM potassium phosphate
pH	9	8
Temperature (K)	277	277
Precipitant	24% (w/v) MPD	30% (w/v) MPD
Additives	pdTp, Calcium Chloride	pdTp, Calcium Chloride
Data collection		
Wavelength (Å)	1.1	0.9790
Resolution (Å)	50.00-1.73 (1.76-1.73)	50.00-1.55 (1.58-1.55)
Unique reflections	14929 (754)	20413 (1005)
Completeness	0.999 (0.996)	0.999 (0.996)
Redundancy	11.9 (9.5)	12.6 (10.0)
Average $I/\sigma(I)$	19.6 (11.5)	26.2 (13.8)
R_{merge}	0.062 (0.211)	0.034 (0.195)
Wilson B ($Å^2$)	38.3	31.7
Space group	$P4_1$	$P4_1$
Cell dimensions (Å ; °)	a = 47.79 ; α = 90.00 b = 47.79 ; β = 90.00 c = 63.31 ; γ = 90.00	a = 47.64 ; α = 90.00 b = 47.64 ; β = 90.00 c = 62.81 ; γ = 90.00
Refinement		
Resolution (Å)	47.79-1.73 (1.78-1.73)	47.64-1.55 (1.59-1.55)
No. of non-hydrogen atoms	1186	1256
No. of unique reflections	14906 (1086)	20389 (1462)
No. of reflections in test set	755 (58)	1045 (73)
R_{work}	0.190 (0.26)	0.177 (0.25)
R_{free}	0.240 (0.29)	0.219 (0.24)
RMSD from ideal geometry		
Bonds (Å)	0.020	0.019
RMS angles (°)	1.99	1.99
Average B-factors ($Å^2$)		
Protein	41.9	27.1
Solvent	37.3	33.9
Ion	34.3	28.0
Ramachandran plot		
Most favored (%)	103 (87.3)	102 (88.7)
Additionally allowed (%)	14 (11.9)	12 (10.4)
Disallowed (%)	1 (0.9)	1 (0.9)
No. of residues excluding Gly, Pro and termini	118	115
Total no. of residues	135	132
PDB accession code	4ODG	4KD4
RMSD (Å) from SNase	0.38	0.40
Main chain only	0.21	0.21

Supplemental Table 1 (continued)

Crystallization conditions	V23I/V66I/I72V/I92V
Buffer	25mM potassium phosphate
pH	9
Temperature (K)	277
Precipitant	20% (w/v) MPD
Additives	pdTp, Calcium Chloride
Data collection	
Wavelength (Å)	0.979
Resolution (Å)	50.00-1.65 (1.68-1.65)
Unique reflections	16933 (826)
Completeness	1.000 (1.000)
Redundancy	13.0 (13.1)
Average $I/\sigma(I)$	20.8 (12.1)
R_{merge}	0.041 (0.274)
Wilson B ($Å^2$)	36.6
Space group	$P4_1$
Cell dimensions (Å ; °)	a = 47.61 ; α = 90.00 b = 47.61 ; β = 90.00 c = 62.91 ; γ = 90.00
Refinement	
Resolution (Å)	47.60-1.65 (1.79-1.65)
No. of non-hydrogen atoms	1190
No. of unique reflections	16911 (1209)
No. of reflections in test set	852 (59)
R_{work}	0.196 (0.28)
R_{free}	0.239 (0.39)
RMSD from ideal geometry	
Bonds (Å)	0.018
RMS angles (°)	1.89
Average B-factors ($Å^2$)	
Protein	36.5
Solvent	36.9
Ion	36.8
Ramachandran plot	
Most favored (%)	99 (86.1)
Additionally allowed (%)	15 (13.0)
Disallowed (%)	1 (0.9)
No. of residues excluding Gly, Pro and termini	115
Total no. of residues	132
PDB accession code	4R8N
RMSD (Å) from SNase	0.39
Main chain only	0.20

[a] Values in parenthesis correspond to the highest resolution shell.

Supplemental figure 1: Normalized pressure unfolding profiles measured by following the center of spectral mass of Trp-140 fluorescence as a function of pressure. Variant V23M in 0.30 M GdnHCl in filled circles (•), V23M/T62F in 0.15 M GdnHCl in crosses (+), WT SNase in 0.4 M GdnHCl in open squares (□), and V23I/V66I/V74I/V99I in 0.30 M GdnHCl in filled triangles (▼).



Supplemental figure 1

CHAPTER 4:

ROLE OF INTERNAL HYDRATION ON THE BEHAVIOR OF PROTEINS UNDER PRESSURE

Abstract

Cavities that are present in the folded state and absent in the unfolded state determine the thermodynamics of pressure unfolding of proteins. In a systematic study of pressure unfolding of variants of staphylococcal nuclease (SNase) with artificial cavities engineered by substitution of core positions with Ala, the cavities appear to be dry in crystal structures. This is consistent with the near-linear relationship found between the volume of cavities observed in the crystal structures and the volume measured by pressure unfolding (ΔV). One conjecture about the mechanism of pressure unfolding invokes pressure-induced water penetration into these cavities that leads to a state of lower volume concomitant with the disruption of interactions that stabilize the folded state. To investigate the effect of internal hydration on ΔV, the volume and polarity of internal cavities were modulated in variants of SNase through polar substitutions at internal positions. Crystal structures were solved in 1 bar pressure to inspect the state of hydration of the cavities. Pressure unfolding experiments monitored by Trp fluorescence were performed to measure ΔV. Proteins with internal polar residues reported both high and low ΔV values that cannot be explained by a simple contribution from electrostriction. The proteins for which the structure showed an artificial cavity fully occupied by internal water molecules still reported increases in ΔV relative to the parent protein that were comparable to those of a protein with a dry cavity of similar volume. This can be reconciled by assuming that the measured ΔV values already account for the effect of internal hydration and that cavities experience partial hydration even in the absence of internal polar groups. The high ΔV values measured are not consistent with

water penetration induced by pressure and can be explained by the structural and volumetric fluctuations that are coupled with the hydration of the protein interior.

Introduction

The hydrophobic interior of proteins often contain internal cavities large enough to fit water molecules and are thought to be central to allostery, catalytic activity (1–3) and to the evolvability of biological function (4–6). Some of these cavities are seen to contain internal water molecules (7) but these are often well hydrogen bonded to protein and considered part of the protein structure (8, 9). Mostly hydrophobic cavities, on the other hand, appear empty in crystal structures (10, 11) and hydrating them would represent, in principle, a large energetic penalty (12). Access and shuttling of waters through protein cores is central to fundamental biological processes like the proton-translocation mechanism in bacteriorhodopsin (13) or the balance of osmotic pressure by aquaporin (14). The state of hydration of protein cavities is difficult to characterize and the determinants of internal hydration are not well understood. Technical innovations in NMR spectroscopy have allowed to map the hydration layer of ubiquitin (Ub) (15, 16) but the protein interior remained dry, consistent with previous findings (17). However, millisecond long simulations (18) have identified water wires and transient tunnels that are consistent with the hydration dynamics observed with relaxation dispersion experiments (19). Pressure experiments suggested that water penetration into the cavity is favored at high pressures (20), a proposed mechanism of pressure unfolding (21). Recently, high pressure NMR spectroscopy coupled with confinement in a reverse micelle has been used to probe the hydration state of the engineered cavity in variant L99A of T4 lysozyme (22). The work indicated that water does access the hydrophobic interior of the protein, but does so only at pressures beyond those needed to unfold the protein.

It has been proposed that the volume of internal cavities is responsible for the difference in volume, ΔV, between the folded and unfolded states that drives the pressure unfolding of proteins (11, 23). A systematic study of SNase variants with cavity generating substitutions to Ala in the protein interior showed a near-linear relationship between the volume of cavities observed in the crystal structures and the ΔV measured by pressure unfolding (11). These cavities appeared to be dry in crystal structures. However, any water molecules buried in internal cavities, too dynamic to resolve by crystallography, could stabilize the protein against high pressure by effectively reducing the volume of cavities (20, 21, 24).

Buried polar residues can also increase the ΔV of the protein through an electrostriction effect (25–27). This refers to the decrease in volume of the solvating water of a polar group upon protein unfolding. This phenomenon has been proposed as a significant factor in determining the behavior of proteins under pressure (28–32). However, pressure perturbation experiments (33, 34) and pressure unfolding experiments (11, 25, 35) argue against a primary role of electrostriction in determining ΔV. Most of these experiments evaluate the differential solvation of the total protein surface (polar and non-polar, backbone and side chains) (35) or evaluate the effect of surface substitutions (34). The results from internal polar substitutions, on the other hand, show that burying a polar group can result in a significant increase in the ΔV of a protein through electrostriction effects (26).

Previous studies on staphylococcal nuclease (SNase) identified discrete sites for water localization in the crystal structures of variants engineered with internal polar groups obtained at cryogenic temperatures (8). Most of these sites, however, did not

display any evidence of hydration in the room temperature crystal structures of the same variants. This is consistent with a reduction in the side chain dynamics induced by cryocooling that leads to a more structured protein interior and therefore the detection of internal water molecules (36–39). Molecular dynamics simulations of the same variants corroborated the increased hydration of the protein interior due to the buried polar side chains and identified specific routes of access from bulk water (40–43). The structural fluctuations that result in transient cavities and transient access to cavities are volumetric fluctuations that can contribute to ΔV.

In this work, the consequences of the hydration of internal cavities on the pressure unfolding of proteins was probed by modulating the state of hydration of internal cavities using polar substitutions near internal cavities. Starting with the previously studied variants of SNase with artificial cavities (11), a set of variants of SNase with buried polar groups near artificial cavities were engineered to evaluate systematically the effect of increased hydration on the pressure sensitivity of proteins. The cavities generated with truncations to Ala were lined with polar side chains by substituting a Ser, Thr, Asn or Gln residue at an internal position. This set included a subset of variants engineered using the isosteric substitution Val-to-Thr that changes the polarity but does no alter the volume of the cavity. Thermodynamic stabilities ($\Delta G°$) were obtained by chemical denaturation at 1 bar. Crystal structures were solved to study the differences in internal hydration patterns. Pressure unfolding monitored by fluorescence yielded ΔV values of the variants. This systematic study provides insight into the structural determinants of internal hydration and its effect on the pressure sensitivity of proteins. It also enables an evaluation of the role of electrostriction in the pressure unfolding of proteins, and provides insight into the

relationship between cavities and internal polar residues and the structural and volumetric fluctuations that are coupled to the hydration of the protein interior.

Results

The 27 proteins studied here comprised the following variants of the highly stable background of SNase, Δ+PHS: V23S, V23T, V23N, V23Q, V66T, I92S, I92N, I92Q, V99T, V23S/V66A, V23N/V66A, V23Q/V66A, V23T/V66T, V23T/V99T, V66A/I92S, V66A/I92N, V66A/I92Q, V66T/V99T, V23T/V66T/V99T, V23T/L25A/V66T, V23T/L25A/V99T, L25A/V66T/V99T, V23T/L36A/V66T, V23T/L36A/V99T, L36A/V66T/V99T, V23T/V66A/V99T, V23A/V66T/V99T and V23T/V66T/V74T/V99T. These variants were selected based on several factors. First, the insight from previous crystallographic studies allowed educated guesses on which substitutions would yield a pattern of internal hydration (8). Positions 23 and 92 in the β-barrel were selected to probe the location dependence of internal hydration. Position 92 is one of the most deeply buried positions in SNase, and substitutions at this site often result in high energetic penalties due to its highly hydrophobic microenvironment, but also in extended networks of internal waters (44–46). In contrast, position 23 is directly adjacent to the cavity observed in SNase, is proximal to the surface of the protein, and can hydrogen bond to an interfacial water molecule observed in the parent protein (47). The polar substitutions spanned residues of all sizes, ranging from Ser to Gln, to probe how the size of the polar group and the degrees of freedom of the side chain influence the extent of hydration. The isosteric substitutions involved the three Val residues that directly line the cavity of SNase, namely Val-23, Val-66 and Val-99. The Ala substitutions V23A, L25A, L36A and V66A were selected from the previous set of 10 variants of SNase with an Ala substitution (11) because these were sites that extended the

cavity in SNase and resulted in a range of ΔV values (from 62 mL/mol to 104 mL/mol) (Table 2).

1. Crystal structures. Crystal structures at cryogenic temperatures and 1 atm of pressure were obtained for 14 variants of staphylococcal nuclease (SNase): V23S, I92S, I92N, I92Q, V23S/V66A, V23Q/V66A, V66A/I92S, V66A/I92N, V66A/I92Q, V23T/V66T, V23T/V99T, V66T/V99T, V23T/L25A/V99T, V23T/L36A/V99T, V23T/V66A/V99T, and L36A/V66T/V99T. The structures were inspected for the presence of water molecules in the vicinity of the substitutions. This inspection was always done with reference to the structure of Δ+PHS variant of SNase. For simplicity, the previously used nomenclature will be used to identify internal waters according to their location (8) if the center of the waters do not deviate more than 1 Å from the site they are labeled with (Fig. 3). Sites 1 and 7 are occupied by waters in the Δ+PHS parent protein; therefore, these water molecules will not be mentioned unless they are absent in the variant structure.

1.1 Bias introduced by cryo-cooling. The structures studied here were solved at temperatures of 100 K, artifacts of which can result in an increase in water structure, compaction of the protein structure and reduction in the dynamics of side chains (8, 36–39, 48). In particular, cryo-cooling results in an increase in observed internal water molecules for buried polar groups with respect to room temperature crystal structures (8).

However, even in crystals at room temperature, evidence is found for waters with partial occupancy at some of the same sites as those observed at cryogenic temperatures (8). Increased kinetic energy at higher temperatures appears to delocalize the buried

water molecules and render them undetectable by crystallography. Molecular dynamics simulations of variants of SNase with internal polar groups show extensive patterns of internal hydration, including those observed by cryo-crystallography (40). Although variants such as I92N (Fig. 2D) do not show internal hydration patterns in the cryogenic structures presented here, it is probable that this is due to the limitations and artifacts of cryo-crystallography (8, 48) and that these variants experience an increase in internal hydration relative to the non-polar interior of the parent protein.

1.2 I92S, I92N and I92Q. In the structures of variants with I92S, I92N and I92Q the polar residue at position 92 is buried in the core of the β-barrel (Fig 2C, 2D and 2E). In the structure of I92S and I92N, the polar side chain interacts with the backbone carbonyl of Glu-73 and no new waters are observed (interfacial waters 1 and 7 are present) (Suppl. Fig. 2C and 2D). Asn-92 shows an alternate conformation with 40% occupancy making a non-optimal contact with the carbonyl of Tyr-91 (Suppl. Fig. 2D). There are no other polar groups within 4 Å of Ser-92 nor within either conformer of Asn-92. In the case of I92Q, the side chain of Gln-92 makes a single polar contact in a radius of 4 Å with a water buried fully at position 6 (Wat-321) (Suppl. Fig. 2E). Because Wat-321 is a hydrogen bond donor to the carbonyls of Arg-35 and Ala-90, it is assumed to be a hydrogen bond acceptor to the amine group of the Gln-92 side chain.

1.3 V66A/I92S, V66A/I92N and V66A/I92Q. The V66A substitution enlarges the internal cavity. The structure of the variant with V66A/I92S, shows a core devoid of water molecules (Fig. 2K) and instead shows how Ser-92 satisfies one hydrogen bond

with the carbonyl oxygen of Glu-73, as observed in the structure of I92S (Suppl. Fig. 2C). In stark contrast with the dry cavity observed in the structure of the I92N variant, the crystal structure of the V66A/I92N variant shows that Asn-92 interacts with five new water molecules buried at sites 2 (Wat-402), 3 (Wat-404), 4 (Wat-403), 5 (Wat-424) and 6 (Wat-423) (Fig. 1L). In the structure of the V66A/I92N variant the side chain of Asn-92 adopts the same two conformers seen in the structure of the I92N variant and with comparable occupancy (Suppl. Fig. 2D). In the first conformer the side chain contacts the water at site 4, which hydrogen bonds to the waters in sites 2 and 3, all of which have full occupancies. Those water molecules are themselves coordinated by the polar backbone of residues 19, 20 and 62 and by the hydroxyl of Thr-62. The second conformer interacts with two water molecules at sites 5 and 6, which have occupancies of 50% that match the occupancy of the Asn-92 conformer. They are also coordinated by the carbonyl oxygen atoms of residues 90, 35 and 34. The second conformer of Asn-92 and its hydration pattern is identical to that observed for Asp-92 in the previously published structure of the I92D variant (8).

In V66A/I92Q, Gln-92 adopts a different conformer than in I92Q and interacts with 3 new water molecules at sites 2 (Wat-369), 3 (Wat-339) and 4 (Wat-366), while site 6 is now devoid of water (Fig. 1M). Because Gln-92 only forms one hydrogen bond to the water at site 4 (Suppl. Fig. 2F), the orientation of its polar group is under-determined and was modeled to hydrogen bond to Wat-67 with its carbonyl oxygen, as suggested by the Molprobity server (49). This conformer and hydration pattern is identical to the one observed with Glu-92 in the previously published structure I92E (8).

The differences and similarities between the structures of SNase variants with internal ionizable residues (I92D and I92E) (8) and those with internal polar residues presented here are remarkable. The determinants of the conformation of internal polar and ionizable residues remain poorly understood (see Thesis Appendix). The comparison enabled by the structures presented in this work highlight the difference in hydration properties of the polar Asn and Gln relative to Asp and Glu. In particular, the conformer of Gln-92 (Fig. 2E) in the structure of variant I92Q is not populated in the structure of I92E, solved at a pH that populates the uncharged species of Glu-92. Instead, the substitution V66A, which incurs an energetic penalty of 2.9 kcal/mol, is necessary to populate the state observed in I92E. This seems to reflect a difference in the fluctuations of side chain conformations and hydration the protein interior induced by ionizable and polar side chains and that relate to the dielectric responses of the protein interior that result in charge stabilization (44–47).

1.4 V23S, V23S/V66A and V23Q/V66A. For comparison, the polar substitution was introduced at position 23, across the β-barrel from position 92 and closer to the protein-water interface. In the structure of the V23S/V66A variant, Ser-23 interacted with a water molecule at site 3 (Wat-309) (Fig. 2I). In the structure of variant V23S, Wat-350 occupied site 3 as well (Fig. 2B), which was coupled to the alternative rotameric state observed for Val-66 that was absent in Δ+PHS and which generated the space for water site 3 (Suppl. Fig. 2A). Hydration of site 3 was coupled with the slight (~5 Å) deviation of β-stand-1 away from α-helix-1, similar to what was observed for variant V23T/V99T (Fig. 5).

The structure of the V23Q/V66A variant showed the side chain of Gln-23 pointing toward position 92 (Fig. 2J). The orientation of the polar group was underdetermined and modeled as suggested by the MolProbity server (49). Gln-23 was hydrogen bonded to a water molecule at site 3 (Wat-333) (Suppl. Fig. 2B) and there was evidence for a water molecule at site 5 and, when modeled in, would report adequate B-factors (~30 $Å^2$) at half occupancy without any negative difference density. However, the electron density for this water molecule was not clear enough nor spherical enough to be convincing and did not remain in the final model. Once again, hydration of water site 3 is coupled with the slight deviation of β-stand-1 away from α-helix-1, as shown in Fig. 5.

1.5 V23T/V66T, V23T/V99T and V66T/V99T. Valine residues at positions 23, 66 and 99 are adjacent to the cavity observed in the structure of the Δ+PHS variant of SNase. The isosteric substitution Val-to-Thr should change the polarity near the cavity without changing the volume of the cavity.

The crystal structure of the V23T/V99T variant showed an empty cavity that is similar to the cavity in Δ+PHS (Fig. 1A and 1G). Volume calculations showed a small increase in cavity volume relative to the parent protein (Table 3). Thr-99 was found in two conformers, both able to hydrogen bond with the carbonyl of Ile-92 (Suppl. Fig. 2H). Thr-23 adopted a conformer equivalent to that of Val-23 as observed in the structure for Δ+PHS nuclease. Thr-23 hydrogen bonds to a water molecule at site 3 (Wat-341) and 8 (Wat-380). Val-66 adopted a different rotameric state than the one observed in the structure of Δ+PHS SNase. Furthermore, the V23T/V99T structure showed a displacement of 0.5 Å of the C-α of Ala-17 and Glu-67 relative to the structure of

Δ+PHS, which locally moves β-strand-1 away from α-helix-1 (Fig. 5). This is coupled to the hydration of sites 3 and 8 as well as an alternative rotameric state observed for Glu-67 and explains the slight increase in calculated cavity volume (Table 3).

In the structure of the V23T/V66T protein Thr-23 was in the same conformer as in the structure of the V23T/V99T protein (Fig. 2F). A water molecule was also found at site 3 (Wat-315) forming hydrogen bonds of 2.7 Å length to both Thr-23 and Thr-66 (Suppl. Fig. 2G). Thr-66 adopted a rotamer equivalent to that of Val-66 as observed in Δ+PHS. No evidence was found for water at site 8 nor for a displacement between the β-strand-1 and α-helix-1. No change in cavity volume was found using a probe of 1.4 Å and the CASTp algorithm, although this depended on the probe size and algorithm used, indicating that the change in cavity volume is at the limit of what the algorithms can detect (Table 3).

In contrast to structures V23T/V66T and V23T/V99T, the structure of the V66T/V99T protein showed the C-α of Ala-17 shifted towards Thr-66 by 0.5 Å compared to Δ+PHS and by 1 Å relative to V23T/V99T (Fig. 5). No water molecules were observed at sites 1, 3 and 8, indicating that this movement occludes water sites 1, 3 and 8, although faint evidence in both $2F_o$-F_c and F_o-F_c maps can be found for a water molecule with partial occupancy at site 1. Thr-66 adopted the same conformer as Val-66 in Δ+PHS and only contacted the carbonyl oxygen of positions 62 and 63 (Suppl. Fig. 2I). Thr-99 adopted the same two conformers that hydrogen bond to the carbonyl of Ile-92 observed in the structure of V23T/V99T. Cavity volume calculations showed a clear decrease in volume for this variant, indicating that the interaction between Thr-66 and the backbone of α-helix-1 is coupled to improved packing.

1.6 V23T/V66A/V99T, V23T/L25A/V99T, V23T/L36A/V99T and L36A/V66T/V99T. Variant V23T/V66A/V99T showed Thr-23 and Thr-99 in the same rotameric states as those observed in V23T/V99T. The truncation to Ala-66 results in the localization of a water molecule at site 3 (Wat-313) that was hydrogen bonded to Thr-23 (Fig. 2P) and no backbone deviations were observed in β-strand-1 relative to the parent protein. The cavity volume calculated from this structure was comparable to that of the structure of variant V66A (Table 3).

The cavity observed in the structure with V23T/V99T was enlarged by introducing the L25A substitution (Fig. 2N). In this V23T/L25A/V99T variant both Thr-23 and Thr-99 adopted multiple conformers. Thr-99 adopted conformers equivalent to those it occupies in the structure of the V23T/V99T variant and it hydrogen bonded to the carbonyl of Ile-92. The main conformer of Thr-23 matched that of found in the structure of the V23T/V99T variant and it contacts a water molecule at site 3 (Wat-443). The secondary conformer formed a less optimal contact with the water at site 3 and with the carbonyl of Thr-22. No backbone deviation of β-strand-1 was observed. The calculated cavity volume was significantly larger than that of L25A, due to a displacement of β-strand-1 in the structure of L25A toward α-helix-1.

The structure of the V23T/L36A/V99T variant showed a hydrated cavity with two full occupancy water molecules (Wat-314 and Wat-371) at sites not identified previously and not represented in Fig. 1. The water molecules form a bridge between the hydroxyls of the Thr residues (Fig. 1O). Each Thr side chains forms two hydrogen bonds (Suppl. Fig. 2J). Thr-99, in a single conformation, contacts the carbonyl of Ile-92 and Wat-314.

Thr-23 also adopts a single conformation, forms hydrogen bonds to Wat-371 and a water in site 1 (Wat-303) and points the hydroxyl group toward the cavity and away from Val-66, which retains the conformation observed in Δ+PHS and occludes water sites 3 and 8. Were Leu-36 present, it would have a steric clash with Wat-314. No backbone deviation of β-strand-1 is observed and the calculated cavity volume is identical to the volume of cavities calculated using structure L36A (Table 3).

In stark contrast with variant V23T/L36A/V99T, the structure of the variant with L36A/V66T/V99T showed a cavity devoid of waters (Fig. 1Q). Thr-99 again formed a hydrogen bond with the carbonyl oxygen of Ile-92 although it is in a different conformer from that observed in the structure of the V23T/L36A/V99T variant. Thr-66 formed hydrogen bonds with the carbonyl of residues 62 and 63. β-strand-1 is locally displaced toward α-helix-1 and occludes water sites 1, 3 and 8, with Ala-17 moving 0.5 Å compared to Δ+PHS. Like in the structure of V66T/V99T, there is evidence of a partially occupied water at site 1 that was not convincing enough to model. The calculated cavity volume is comparable to that of the structure of L36A (Table 3).

2. Cavity volume calculations. Cavity detection algorithms typically use a probe that approximates the size of a water molecule to define the protein surface (50–52). The probe is placed over the entire protein and the contact points between the probe and the protein atoms define the surface of the protein (53). The size of the probe and the volume calculation method used by the algorithm will influence the number and volume of cavities found. The cavities in the crystal structures studied here were studied by comparing volumes obtained by probes of 1.2 and 1.4 Å and by two different algorithms,

CASTp (52) and McVol (51), which use a Delaunay and a Montecarlo approach, respectively. All calculations were visually cross-checked with the crystal structure to confirm that the cavities detected were near the substitution sites and only cavities near the substitution site were considered (Fig. 2). This was done because algorithms that calculate volumes from structure are limited by the use of a probe with a specific, user-determined radius. For example, a probe of radius 1.4 Å will not detect a spherical volume of radius 1.39 Å even though it is only 0.2 mL/mol smaller than a spherical volume of 1.4 Å. Very small changes in the structure that might lead to unrealistic changes in the volume of cavities are a typical result from crystallization and can be observed when comparing two crystals of the same protein, or even when comparing two copies of the same protein within the same asymmetric unit. These small structural changes can therefore lead to misleading differences in the volume of cavities.

The presence of internal water molecules can reduce significantly the void volume detected in the crystal structure of a protein. To evaluate the effect of internal waters on the calculated volume of cavities, calculations were run in the presence and absence of any internal water molecules identified in the crystal structures (Fig. 2). In the case of variant V23T/L36A/V99T, including the water molecules in the calculation resulted in no detectable cavity with a probe of 1.4 Å and using the McVol algorithm (Table 3). Calculations with a probe of 1.4 Å and using the McVol algorithm of the structures of variants I92Q and V23Q/V66A reported no cavities in the presence and absence of internal water molecules. This is consistent with a longer Gln side chain that projects into the original cavity and divides it into smaller cavities that are undetectable to

a probe of 1.4 Å. A smaller probe or the CASTp algorithm is capable of detecting cavity volumes in these variants.

3. Thermodynamic stability. The thermodynamic stability (ΔG°) of variants with internal polar groups was measured by GdnHCl titration monitored with the fluorescence of Trp-140 (Fig. 3 and Table 1). The energetic penalty for polar substitutions at position 92 ranged from 6.5 to 7.7 kcal/mol for I92S and I92Q respectively. Polar substitutions at position 23, on the other hand, resulted in energetic penalties ranging from 3.4 to 4.9 kcal/mol, for V23T and V23N, respectively. This reflects the differences in microenvironment of positions 23 and 92. Position 92 is located in a hydrophobic environment far from the protein surface, whereas position 23 can readily interact with the interfacial water at site 1 and faces the cavity, which can accommodate large side chains or even internal waters. In general, substitution to Asn was always the more costly polar substitution. Asn is known to be a structure breaker that interacts well with its own backbone because of its short side chain and large polar group (54).

Variants with one, two, three and four Val-to-Thr substitutions were increasingly less stable in a non-additive fashion. Variants with two Thr substitutions are generally more stable than expected from the stability of the single variants by up to 3.6 kcal/mol. The V23T/V99T variant, however, is 1.6 kcal/mol less stable than predicted from additivity consistent with a decrease in packing efficiency of the V23T/V99T variant due to the displacement of β-strand-1 (Fig. 5). When substitution V66T is introduced, the triple variant V23T/V66T/V99T is once again 3.3 kcal/mol more stable than expected from additivity. Variant V23T/V66T/V74T/V99T was the least stable with 0.5 kcal/mol

stability but still 1.4 kcal/mol more stable than expected from additivity. This suggests that the main energetic penalty is related to the conformational rearrangement induced by V23T/V99T.

When the cavity around the buried polar groups at position 23 and 92 is enlarged using the V66A substitution, the resulting double variant is always more stable than expected from simple addition of the energetic cost of the single variants. The polar side chain with the longest aliphatic chain, Gln, deviates from additivity the most: V66A/I92Q is 1.3 kcal/mol more stable than expected (Table 1). The large Gln side chain is able to rearrange and relocate closer to the cavity generated by the V66A substitution and find a more optimal hydration network, as shown in the structure of V66A/I92Q (Fig. 1M).

For variant V23T/V99T, introducing a cavity at positions 25, 36 or 66 always resulted in variants that were up to 1 kcal/mol more stable than expected from additivity (Table 1). This is consistent with the idea that the largest energetic penalty has already been paid for by the V23T/V99T substitutions that induce the structural rearrangement observed in Fig. 5. Similarly, introducing a cavity in variant V66T/V99T at positions 23 and 25 resulted in slightly more stable variants than expected from additivity. Variant L36A/V66T/V99T as well as variants V23T/L25A/V66T and V23T/L36A/V66T reported stabilities within 0.5 kcal/mol from the value expected from additivity.

The *m*-values of variants with one or more polar substitutions ranged from 4.7 to 5.8 kcal mol^{-1} M^{-1} (Table 1). In general, the polar substitutions increased the *m*-value of SNase to up to 5.8 kcal $mol^{-1}M^{-1}$ as in the case of V23T/V66T/V99T. Enlarging the cavity volume by introducing a substitution to Ala almost always resulted in an increase in *m*-value, with the largest values corresponding to V66A/I92N and V23T/L36A/V99T

of up to 1.6 kcal $mol^{-1}M^{-1}$. The increases in *m*-values are consistent with increased association with the denaturant in the unfolded state due to increased polar surface area, either due to the polar side chains or to the more exposed backbone upon truncation of the side chain to Ala (55).

4. Pressure unfolding monitored by Trp fluorescence. The pressure unfolding of a subset of SNase variants was monitored by fluorescence of Trp-140 (Fig. 4 and table 2). The ΔV of Δ+PHS was previously estimated as 58 mL/mol (Table 2) (11). The single variants with I92N or I92Q substitutions report a ΔV value of 84 mL/mol, an increase of 26 and 27 mL/mol compared to Δ+PHS. The increase in ΔV reflects in part the change in van der Waals volume of -17 and -6 mL/mol of the substitution, respectively (Table 3). Variants V23T/V99T, V23T/V66T/V99T and V23T/V66T/V74T/V99T also reported an increase in ΔV of 27 and 19 mL/mol, respectively (Table 2). These variant contained isosteric Val-to-Thr substitutions, which did not significantly alter the volume of the cavity (Table 3). The changes in ΔV for the 4 variants I92N, I92Q, V23T/V99T, V23T/V66T/V99T and V23T/V66T/V74T/V99T appear to be consistent with an electrostriction effect due to the substituted polar groups (26).

However, variant V66T/V99T appeared as an outlier of this trend, reporting a ΔV of 53 mL/mol, comparable to that of Δ+PHS (58 mL/mol) (Table 2). The change in ΔV of -5 mL/mol, the only variant that reported a reduction in ΔV, relative to the parent protein cannot be explained through an electrostriction effect, since variant V23T/V99T reports a much greater ΔV of 85 mL/mol and buries the same two polar side chains. Moreover, the non-additivity of changes in ΔV measured for variants V23T/V66T/V99T

and V23T/V66T/V74T/V99T is also inconsistent with an electrostriction effect (Table 2). The structural deviation observed for this variant (Fig. 5) results in a smaller cavity volume as calculated from the structure (Table 3) relative to the parent protein, which can explain the low ΔV value for this variant.

Proteins that contained an artificial cavity reported no change in ΔV when polar residues are introduced at internal positions near the cavity. Variants V23T/L36A/V66T, V23T/L36A/V99T, L36A/V66T/V99T and V23T/L36A/V66T/V99T reported values within error (ΔV of 107, 96, 103 and 104, respectively) of that of variant L36A, ΔV = 104 mL/mol. Similarly, variants V23T/L25A/V66T, V23T/L25A/V99T, L25A/V66T/V99T and V23T/L25A/V66T/V99T (ΔV of 111, 97, 106 and 113, respectively) reported values comparable to that of variant L25A, ΔV = 96 mL/mol. Variant V23T/V66A/V99T reported a ΔV = 82 mL/mol, comparable to that of variant V66A, ΔV = 70 mL/mol. This is inconsistent with any significant contribution from electrostriction.

Furthermore, when substitutions V66T/V99T are introduced to variant V23A, ΔV = 74 mL/mol, the resulting triple variant V23A/V66T/V99T reports a very large ΔV of 127 mL/mol. This dramatic increase cannot be rationalized by a contribution from electrostriction due to buried polar groups, since variant V66T/V99T reports a low ΔV of 53 mL/mol. The large increase in ΔV of over 50 mL/mol observed for variant V23A/V66T/V99T relative to V66T/V99T is significantly larger than the increase (16 mL/mol) observed for the single variant V23A relative to Δ+PHS. It is possible that the ΔV of this variant originates from the increase in cavity volume due to Ala-23 as well as

increased structural fluctuations that are coupled to the hydration of water sites 1, 3 and 8.

Discussion

Effect of internal polar residues on ΔV. The change in exposed surface area upon unfolding underlies the thermal and chemical unfolding of proteins (55, 56), and is thought to play a significant role in the pressure sensitivity through an electrostriction effect (26, 27). This phenomenon is due to the decrease in volume of the solvating water of an internal polar group upon protein unfolding. However, a lack of correlation between ΔV and the change in exposed surface upon unfolding indicates that electrostriction must not contribute significantly to the pressure sensitivity of proteins (35). Instead, the contribution to ΔV from internal cavities has been proposed as the dominant factor based on a linear relationship established between the ΔV of proteins and the volume of cavities calculated from structure (11)(Caro et al., manuscript in preparation).

Previous studies of variants of SNase with internal polar and ionizable residues reported increases in ΔV values relative to the parent protein (26). Many of the variants involved internal ionizable residues that were expected to become ionized upon unfolding. This should, in principle, represent the largest contribution from electrostriction due to a single residue. Surprisingly, the largest change in ΔV was measured for variant V66N, with the neutral Asn. The study involved variants with polar groups and side chains of various sizes, which convoluted changes in cavity volumes with contributions from electrostriction. Nevertheless, the measured ΔV values appear to argue that the volume of internal cavities is not the only contribution to ΔV and that electrostriction needs to be taken into account.

To resolve the apparent contradiction in the literature, the effect of internal polar residues on the ΔV of proteins was studied in the present work by introducing polar

groups of various sizes (Thr, Asn, Gln) in the interior of SNase and doing so in an incremental fashion to up to four internal polar groups. Variants I92N, I92Q, V23T/V99T, V23T/V66T/V99T and V23T/V66T/V74T/V99T all reported an increase in ΔV in the range of 19-27 mL/mol relative to the parent protein (ΔV = 58 mL/mol) (Table 2). These large values are consistent with the values obtained previously for variants of SNase with one polar or ionizable residue buried in the interior (26). The difference in side chain volume upon substitution to the polar group in variants I92N and I92Q (-17 and -6 mL/mol) (Table 3) must partially contribute to increases in ΔV observed (11). However, variants with isosteric Val-to-Thr substitutions experience minimal changes in side chain volume (Table 3) and still show a comparable increase in ΔV (Table 2).

One variant deviates from this trend. Variant V66T/V99T (ΔV = 53 mL/mol) does not show an increase in ΔV relative to the parent protein Δ+PHS (ΔV = 58 mL/mol) (Table 2). A negligible effect on ΔV is also observed for variants where an internal cavity is lined by polar residues. For example, variant L36A reports a ΔV between 95 and 104 mL/mol (Table 2) (11). The cavity generated by Ala-36 can be lined by iso-steric substitutions V23T, V66T and V99T, in any combination, yielding variants such as V23T/L36A/V66T/V99T. The ΔV values of this variant is 104 mL/mol, identical to that of L36A. Similarly, the cavities generated by substitutions Ala-25 and Ala-66, can be lined by substitutions V23T, V66T and V99T and yield a small or no change relative to the single Ala variants L25A and V66A.

Together with the low ΔV value measured for V66T/V99T, these data are difficult to rationalize in the context of electrostriction. It is possible that the effect of internal polar groups is to increase the types of structural fluctuations (Fig. 5) that can have a

significant effect on the interstitial volumes of protein atoms and thus lead to the changes in ΔV of the protein.

Effect of internal polar groups on the hydration of cavities. Introducing a water molecule in a cavity in a hydrophobic environment such as the interior of proteins is expected to result in a large energetic penalty and be strongly disfavored (12, 57). However, fundamental processes in biology such as proton transport and energy transduction (13, 58) rely on the transport and residence of waters within protein interiors. While some proteins incorporate protein-water interactions in their final tertiary fold (17), for other proteins it is clear that if internal waters are present, their dynamics must be too fast to detect (3, 15). Staphylococcal nuclease displays a small cavity in its interior that appears devoid of waters in the crystal structure (8). Introduction of an internal ionizable residue, however, results in extensive patterns of internal hydration (8) and simulations indicate that specific routes are involved in the hydration of the internal ionizable residue (19, 40, 41, 59). Internal hydration is expected to effectively reduce the volume of internal cavities and has been proposed to be favored at increased pressures (21). The variants studied in the present work offer a unique opportunity to evaluate the effect of internal waters on the ΔV of SNase.

When polar substitutions are introduced to a variant with an artificial cavity, the structures of the resulting variants can show cavities that are filled by internal waters (V66A/I92N, V66A/I92Q and V23T/L36A/V99T) (Fig. 2L, 2M and 2O). Some of the same water sites are observed to be hydrated in crystals at room temperature, indicating that the increased internal hydration is not due to an artifact from cryogenic

crystallography (8). This increase in occupancy of internal waters is expected to effectively decrease the volume of the cavity and therefore its contribution to ΔV. The data presented here, however, show that generating an artificial cavity in a protein with an internal polar group always increases the ΔV of the protein (Table 2). Variants V66A/I92N, V66A/I92Q and V23T/L36A/V99T report large ΔV values that range between 91 and 96 mL/mol, despite extensive hydration of the internal cavities.

A proposed mechanism of pressure unfolding involves the pressure-induced penetration of water into the protein interior (21, 60). Scarce and ambiguous experimental data can be found that seem to support this mechanism (20, 22, 61–64). Recent innovations in NMR spectroscopy in combination with reverse micelle technology allowed the detection of interactions between amide nitrogens near the artificial cavity of T4 lysozyme variant L99A and water molecules that were interpreted to reside in the cavity. However, these interactions were detected only at pressures above the unfolding pressure of the protein, which were accessible due to the suppression of the unfolded state in the confined volume of a reverse micelle (22). The data from the present work is not consistent with the proposed argument that increasing pressure will drive waters into the cavities of the protein and drive pressure unfolding by disruption of protein-protein interactions. Instead, they suggest that internal cavities promote the structural fluctuations that are coupled with internal hydration, and are partially hydrated even in the absence of internal polar residues.

The motion observed for variants V23T/V99T and V66T/V99T (Fig. 5) suggests a mechanism for how structural fluctuations can lead to transient cavities and transient access to cavities. Substitutions that alter the polarity and the cavity volume in the protein

interior must affect the heterogeneous side chain dynamics observed in the interior of proteins by NMR spectroscopy (3, 65, 66) and the broad range of residence times observed for internal waters by magnetic relaxation dispersion (17, 19, 67). Increased structural fluctuations coupled to the hydration of internal polar residues result in formation of transient cavities and transient access to cavities (19, 59) that can explain the striking difference in ΔV values measured for variants V23T/V99T and V66T/V99T, and suggest that small increases in interstitial volumes due to fluctuations outweigh any contribution from electrostriction (25, 27).

Conclusion

This work presents a systematic study of the effect of internal hydration on the pressure sensitivity of proteins. Substitutions that line internal cavities with polar residues were used to hydrate the cavities, as observed in the crystal structures. Variants with internal polar residues reported both increases and decreases in ΔV that cannot be rationalized based on a contribution from electrostriction. Some variants with artificial cavities lined with polar groups displayed well ordered buried water molecules that fully occupied and eliminated the artificial cavity. These variants, however, reported ΔV values that were significantly larger than that of the parent protein. Cavities, whether hydrated or not, led to increases in measured ΔV, suggesting that these are partially hydrated even in the absence of internal polar groups. This is inconsistent with the pressure driven hydration of the protein interior and a consequent disruption of stabilizing interactions. Instead, the data indicate that the structural fluctuations induced by internal cavities and internal polar residues lead to significant increases in ΔV and become unfavorable at increased pressures.

Acknowledgements

This work was designed and performed by José Alfredo Caro under the supervision of his advisor Dr. Bertrand García-Moreno E. (Johns Hopkins University, Baltimore, MD). It was possible thanks to a collaboration with Dr. Catherine Ann Royer (Rensselaer Polytechnic Institute, Troy, NY), who served as co-advisor in the design and execution of the experiments, Dr. Jamie L. Schlessman (U.S. Naval Academy, Annapolis, MD) who was instrumental in the production and collection of protein crystals and in the solution of crystal structures, Dr. Mariano Dellarole and Martin Fossat (Centre de Biochimie Structurale, Montpellier, France), who participated in both collection and analysis of high pressure fluorescence data, and to the crystal structures of variants V23S and I92S, produced by Jaime Sorenson.

Bibliography

1. Lee C, Maeng J, Kocher J, Lee B (2001) Cavities of α1-antitrypsin that play structural and functional roles. *Protein Sci.* 10(7):1446–1453.

2. Wiener R, Zhang X, Wang T, Wolberger C (2012) The mechanism of OTUB1-mediated inhibition of ubiquitination. *Nature* 483(7391):618–622.

3. Fu Y, Kasinath V, Moorman VR, Nucci NV, Hilser VJ, Wand AJ (2012) Coupled motion in proteins revealed by pressure perturbation. *J. Am. Chem. Soc.* 134(20):8543–8550.

4. Tokuriki N, Tawfik DS (2009) Protein Dynamism and Evolvability. *Science* 324(5924):203–207.

5. López CJ, Yang Z, Altenbach C, Hubbell WL (2013) Conformational selection and adaptation to ligand binding in T4 lysozyme cavity mutants. *Proc. Natl. Acad. Sci. U.S.A.* 110(46):E4306–E4315.

6. Bouvignies G, Vallurupalli P, Hansen DF, Correia BE, Lange O, Bah A, Vernon RM, Dahlquist FW, Baker D, Kay LE (2011) Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* 477(7362):111–114.

7. Williams MA, Goodfellow JM, Thornton JM (1994) Buried waters and internal cavities in monomeric proteins. *Protein Sci.* 3(8):1224–35.

8. Schlessman JL, Abe C, Gittis A, Karp DA, Dolan MA, García-Moreno E. B (2008) Crystallographic study of hydration of an internal cavity in engineered proteins with buried polar or ionizable groups. *Biophys. J.* 94(8):3208–16.

9. Sun T, Lin F-H, Campbell RL, Allingham JS, Davies PL (2014) An antifreeze protein folds with an interior network of more than 400 semi-clathrate waters. *Science* 343(6172):795–8.

10. Baase WA, Liu L, Tronrud DE, Matthews BW (2010) Lessons from the lysozyme of phage T4. *Protein Sci.* 19(4):631–41.

11. Roche J, Caro JA, Norberto DR, Barthe P, Roumestand C, Schlessman JL, Garcia AE, García-Moreno E. B, Royer CA (2012) Cavities determine the pressure unfolding of proteins. *Proc. Natl. Acad. Sci. U.S.A.* 109(18):6945–6950.

12. Wolfenden R, Radzicka A (1994) On the probability of finding a water molecule in a nonpolar cavity. *Science* 265(5174):936–937.

13. Lanyi JK (2004) Bacteriorhodopsin. *Annu. Rev. Physiol.* 66:665–88.

14. Murata K, Mitsuoka K, Hirai T, Walz T, Agre P, Heymann JB, Engel A, Fujiyoshi Y (2000) Structural determinants of water permeation through aquaporin-1. 599–605.

15. Nucci NV, Pometun MS, Wand AJ (2011) Site-resolved measurement of water-protein interactions by solution NMR. *Nat. Struct. Mol. Biol.* 18(2):245–249.

16. Nucci NV, Pometun MS, Wand AJ (2011) Mapping the hydration dynamics of ubiquitin. *J. Am. Chem. Soc.* 133(32):12326–12329.

17. Denisov VP, Halle B (1995) Protein hydration dynamics in aqueous solution: a comparison of bovine pancreatic trypsin inhibitor and ubiquitin by oxygen-17 spin relaxation dispersion. *J. Mol. Biol.* 245(5):682–97.

18. Shaw DE, Maragakis P, Lindorff-larsen K, Piana S, Shan Y, Wriggers W (2010) Atomic-Level Characterization of the structural dynamics of proteins. *Science* 330:341–347.

19. Persson F, Halle B (2013) Transient access to the protein interior: simulation versus NMR. *J. Am. Chem. Soc.* 135(23):8735–48.

20. Collins MD, Hummer G, Quillin ML, Matthews BW, Gruner SM (2005) Cooperative water filling of a nonpolar protein cavity observed by high-pressure crystallography and simulation. *Proc. Natl. Acad. Sci. U.S.A.* 102(46):16668–16671.

21. Hummer G, Garde S, García AE, Paulaitis ME, Pratt LR (1998) The pressure dependence of hydrophobic interactions is consistent with the observed pressure denaturation of proteins. *Proc. Natl. Acad. Sci. U.S.A.* 95(4):1552–1555.

22. Nucci NV, Fuglestad B, Athanasoula EA, Wand AJ (2014) Role of cavities and hydration in the pressure unfolding of T4 lysozyme. *Proc. Natl. Acad. Sci. U.S.A.* 111(38):13846–13851.

23. Ando N, Barstow B, Baase WA, Fields A, Matthews BW, Gruner SM (2008) Structural and thermodynamic characterization of T4 lysozyme mutants and the contribution of internal cavities to pressure denaturation. *Biochemistry* 47(42):11097–11109.

24. Day R, Garcia AE (2008) Water penetration in the low and high pressure native states of ubiquitin. *Proteins Struct. Funct. Bioinforma.* 70(4):1175–1184.

25. Frye KJ, Perman CS, Royer CA (1996) Testing the correlation between ΔA and ΔV of protein unfolding using m value mutants of staphylococcal nuclease. *Biochemistry* 35(31):10234–10239.

26. Brun L, Isom DG, Velu P, García-Moreno E. B, Royer CA (2006) Hydration of the Folding Transition State Ensemble of a Protein. *Biochemistry* 45(11):3473–3480.

27. Urbauer JL, Ehrhardt MR, Bieber RJ, Flynn PF, Wand AJ (1996) High-Resolution Triple-Resonance NMR Spectroscopy of a Novel Calmodulin-Peptide Complex at Kilobar Pressures. *J. Am. Chem. Soc.* 118(45):11329–11330.

28. Weber G (1992) *Protein interactions* (New York: Chapman and Hall).

29. Chalikian TV, Macgregor Jr RB (2009) Origins of pressure-induced protein transitions. *J. Mol. Biol.* 394(5):834–842.

30. Rasper J, Kauzmann W (1962) Volume Changes in Protein Reactions. 1. Ionization Reactions of Proteins. *J. Am. Chem. Soc.* 84(10):1771–1777.

31. Kauzmann W, Bodanszky A, Rasper J (1962) Volume Changes in Protein Reactions. 2. Comparison of Ionization Reactions in Proteins and Small Molecules. *J. Am. Chem. Soc.* 84(10):1777–1788.

32. Chalikian TV, Gindikin VS, Breslauer KJ (1998) Hydration of diglycyl tripeptides with non-polar side chains: a volumetric study. *Biophys. Chem.* 75(1):57–71.

33. Royer CA, Winter R (2011) Protein hydration and volumetric properties. *Curr. Opin. Colloid Interface Sci.* 16(6):568–571.

34. Vasilchuk D, Pandharipande PP, Suladze S, Sanchez-Ruiz JM, Makhatadze GI (2014) Molecular determinants of expansivity of native globular proteins: a pressure perturbation calorimetry study. *J. Phys. Chem. B* 118(23):6117–6122.

35. Rouget J-B, Aksel T, Roche J, Saldana J-L, Garcia AE, Barrick D, Royer CA (2011) Size and sequence and the volume change of protein folding. *J. Am. Chem. Soc.* 133(15):6020–6027.

36. Halle B (2004) Protein hydration dynamics in solution: a critical survey. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 359(1448):1207–23; discussion 1223–4, 1323–8.

37. Lang PT, Ng H-L, Fraser JS, Corn JE, Echols N, Sales M, Holton JM, Alber T (2010) Automated electron-density sampling reveals widespread conformational polymorphism in proteins. *Protein Sci.* 19(7):1420–31.

38. Fraser JS, Clarkson MW, Degnan SC, Erion R, Kern D, Alber T (2009) Hidden alternative structures of proline isomerase essential for catalysis. *Nature* 462(7273):669–73.

39. Fraser JS, van den Bedem H, Samelson AJ, Lang PT, Holton JM, Echols N, Alber T (2011) Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci. U.S.A.* 108(39):16247–16252.

40. Damjanović A, Schlessman JL, Fitch CA, García AE, García-Moreno E. B (2007) Role of flexibility and polarity as determinants of the hydration of internal cavities and pockets in proteins. *Biophys. J.* 93(8):2791–804.

41. Damjanović A, García-Moreno E. B, Lattman EE, García AE (2005) Molecular dynamics study of water penetration in staphylococcal nuclease. *Proteins Struct. Funct. Bioinforma.* 60(3):433–49.

42. Damjanović A, García-Moreno E. B, Lattman EE, García AE (2005) Molecular dynamics study of hydration of the protein interior. *Comput. Phys. Commun.* 169(1-3):126–129.

43. Damjanović A, Brooks BR, García-Moreno E. B (2011) Conformational Relaxation and Water Penetration Coupled to Ionization of Internal Groups in Proteins. *J. Phys. Chem.* 115:4042–4053.

44. Isom DG, Castañeda CA, Cannon BR, Velu P, García-Moreno E. B (2010) Charges in the hydrophobic interior of proteins. *Proc. Natl. Acad. Sci. U.S.A.* 107(37):16096–16100.

45. Isom DG, Cannon BR, Castañeda CA, Robinson A, García-Moreno E. B (2008) High tolerance for ionizable residues in the hydrophobic interior of proteins. *Proc. Natl. Acad. Sci. U.S.A.* 105(46):17784–8.

46. Isom DG, Castañeda CA, Cannon BR, García-Moreno E. B (2011) Large shifts in pK a values of lysine residues buried inside a protein. *Proc. Natl. Acad. Sci. U.S.A.* 108(13):5260–5265.

47. Robinson AC, Castañeda CA, Schlessman JL, García-Moreno E. B (2014) Structural and thermodynamic consequences of burial of an artificial ion pair in the hydrophobic interior of a protein. *Proc. Natl. Acad. Sci. U.S.A.*

48. Halle B (2004) Biomolecular cryocrystallography: structural changes during flash-cooling. *Proc. Natl. Acad. Sci. U.S.A.* 101(14):4793–8.

49. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D - Biol. Crystallogr.* 66(1):12–21.

50. Kleywegt GJ, Jones TA (1994) Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr. Sect. D - Biol. Crystallogr.* 50(2):178–185.

51. Till MS, Ullmann GM (2010) McVol - a program for calculating protein volumes and identifying cavities by a Monte Carlo algorithm. *J. Mol. Model.* 16(3):419–429.

52. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* 34(suppl 2):W116–W118.

53. Richards FM (1977) Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* 6:151–176.

54. Nick Pace C, Martin Scholtz J (1998) A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. *Biophys. J.* 75(1):422–427.

55. Shortle D (1995) Staphylococcal nuclease: a showcase of m-value effects. *Adv. Protein Chem.* 46:217–247.

56. Kauzmann W (1959) Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* 14:1–63.

57. Rasaiah JC, Garde S, Hummer G (2008) Water in nonpolar confinement: from nanotubes to proteins and beyond. *Annu. Rev. Phys. Chem.* 59:713–40.

58. Steed PR, Fillingame RH (2009) Aqueous accessibility to the transmembrane regions of subunit c of the Escherichia coli F1F0 ATP synthase. *J. Biol. Chem.* 284(35):23243–50.

59. Denisov VP, Schlessman JL, García-Moreno E. B, Halle B (2004) Stabilization of internal charges in a protein: water penetration or conformational change? *Biophys. J.* 87(December):3982–3994.

60. Sarupria S, Ghosh T, García AE, Garde S (2010) Studying pressure denaturation of a protein by molecular dynamics simulations. *Proteins Struct. Funct. Bioinforma.* 78(7):1641–51.

61. Nagae T, Kawamura T, Chavas LMG, Niwa K, Hasegawa M, Kato C, Watanabe N (2012) High-pressure-induced water penetration into 3-isopropylmalate dehydrogenase. *Acta Crystallogr. Sect. D - Biol. Crystallogr.* 68(3):300–309.

62. Liu L, Quillin ML, Matthews BW (2008) Use of experimental crystallographic phases to examine the hydration of polar and nonpolar cavities in T4 lysozyme. *Proc. Natl. Acad. Sci. U.S.A.* 105(38):14406–11.

63. Ernst JA, Clubb RT, Zhou H-X, Gronenborn AM, Clore GM (1995) Demonstration of Positionally Disordered Water Within a Protein Hydrophobic Cavity by NMR. *Science* 267(5205):1813–1817.

64. Matthews BW, Morton AG, Dahlquist FW, Ernst JA, Clubb RT, Zhou H-X, Gronenborn AM, Clore GM (1995) Use of NMR to Detect Water Within Nonpolar Protein Cavities. *Science* 270(5243):1847–1849.

65. Lee AL, Sharp KA, Kranz JK, Song X-J, Wand AJ (2002) Temperature dependence of the internal dynamics of a calmodulin-peptide complex. *Biochemistry* 41(46):13814–13825.

66. Igumenova TI, Frederick KK, Wand AJ (2006) Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution. *Chem. Rev.* 106(5):1672–99.

67. Denisov VP, Halle B (1995) Hydrogen exchange and protein hydration: the deuteron spin relaxation dispersions of bovine pancreatic trypsin inhibitor and ubiquitin. *J. Mol. Biol.* 245(5):698–709.

68. Castañeda CA, Fitch CA, Majumdar A, Khangulov V, Schlessman JL, García-Moreno E. B (2009) Molecular determinants of the pKa values of Asp and Glu residues in staphylococcal nuclease. *Proteins Struct. Funct. Bioinforma.* 77(3):570–88.

69. The PyMOL Molecular Graphics System, Version 1.3 Schrödinger, LLC.

Table 1: Thermodynamic values from chemical denaturation experiments[a, b]

Protein	ΔG^{o}_{H2O} (kcal/mol)	$\Delta\Delta G^{o}_{H2O}$ (kcal/mol)	$\Delta\Delta G^{o}_{Expected}$ (kcal/mol)	$\Delta\Delta G^{o}_{Measured}$ - $\Delta\Delta G^{o}_{Expected}$ (kcal/mol)	*m*-value (kcal/mol*M)	C_{mid} (M)
Δ+PHS	11.9 (0.1)	--	--	--	4.9 (0.1)	2.4 (0.1)
V23A	8.7 (0.1)	-3.2 (0.1)	--	--	5.2 (0.1)	1.7 (0.1)
V23S^	7.4 (0.1)	-4.5 (0.1)	--	--	5.5 (0.1)	1.3 (0.1)
V23T^	8.5 (0.2)	-3.4 (0.2)	--	--	5.2 (0.1)	1.6 (0.1)
V23N*	7.0 (0.1)	-4.9 (0.1)	--	--	5.6 (0.2)	1.2 (0.1)
V23Q*	6.8 (0.1)	-5.1 (0.1)	--	--	5.5 (0.2)	1.2 (0.1)
L25A	8.9 (0.1)	-3.0 (0.1)	--	--	5.0 (0.1)	1.8 (0.1)
L36A	8.6 (0.1)	-3.3 (0.1)	--	--	5.4 (0.1)	1.6 (0.1)
V66A	9.0 (0.2)	-2.9 (0.2)	--	--	4.8 (0.1)	1.9 (0.1)
V66T^	7.7 (0.2)	-4.2 (0.2)	--	--	4.7 (0.1)	1.6 (0.1)
V74T^	7.5 (0.1)	-4.4 (0.1)	--	--	4.6 (0.1)	1.6 (0.1)
I92A	7.9 (0.3)	-4.0 (0.3)	--	--	5.4 (0.2)	1.5 (0.2)
I92S^	5.4 (0.1)	-6.5 (0.1)	--	--	5.7 (0.1)	0.9 (0.1)
I92N*	4.2 (0.1)	-7.7 (0.1)	--	--	5.7 (0.2)	0.7 (0.1)
I92Q*	5.0 (0.1)	-6.9 (0.1)	--	--	5.7 (0.2)	0.9 (0.1)
V99T^	10.1 (0.5)	-1.8 (0.5)	--	--	4.8 (0.3)	2.1 (0.5)
V23S/V66A	5.2 (0.1)	-6.7 (0.1)	-7.4	0.7	5.9 (0.1)	0.9 (0.1)
V23N/V66A	4.7 (0.1)	-7.2 (0.1)	-7.8	0.6	6.1 (0.1)	0.8 (0.1)
V23Q/V66A	5.1 (0.3)	-6.8 (0.3)	-8.0	1.2	5.7 (0.3)	0.9 (0.1)
V23T/V66T	8.0 (0.1)	-3.9 (0.1)	-7.6	3.7	5.8 (0.1)	1.4 (0.1)
V23T/V99T	5.2 (0.1)	-6.7 (0.1)	-5.2	-1.5	5.6 (0.1)	0.9 (0.1)
V66A/I92S	3.1 (0.2)	-8.8 (0.2)	-9.4	0.6	6.0 (0.2)	0.5 (0.1)
V66A/I92N	2.0 (0.2)	-9.9 (0.2)	-10.6	0.7	6.4 (0.2)	0.3 (0.1)
V66A/I92Q	3.4 (0.2)	-8.5 (0.2)	-9.8	1.3	6.3 (0.2)	0.5 (0.1)
V66T/V99T	7.1 (0.2)	-4.8 (0.2)	-6.0	1.2	4.9 (0.2)	1.4 (0.2)
V23T/V66T/V99T	4.4 (0.1)	-7.5 (0.1)	-9.4	1.9	5.8 (0.1)	0.8 (0.1)
V23T/L25A/V66T	4.8 (0.1)	-7.1 (0.1)	-6.9	-0.2	6.0 (0.1)	0.8 (0.1)
V23T/L25A/V99T	2.4 (0.2)	-9.5 (0.2)	-9.7	0.2	6.4 (0.2)	0.4 (0.1)
L25A/V66T/V99T	4.6 (0.2)	-7.3 (0.2)	-7.8	0.5	5.3 (0.1)	0.9 (0.1)
V23T/L36A/V66T	4.3 (0.1)	-7.6 (0.1)	-7.2	-0.4	5.9 (0.1)	0.7 (0.1)
V23T/L36A/V99T	2.7 (0.1)	-9.2 (0.1)	-10.0	0.8	6.5 (0.1)	0.4 (0.1)
L36A/V66T/V99T	3.7 (0.1)	-8.2 (0.1)	-8.1	-0.1	6.0 (0.1)	0.6 (0.1)
V23T/V66A/V99T	3.3 (0.1)	-8.6 (0.1)	-9.6	1.0	6.1 (0.1)	0.5 (0.1)
V23A/V66T/V99T	4.5 (0.3)	-7.4 (0.3)	-8.0	0.6	5.5 (0.3)	0.8 (0.1)
V23T/V66T/V74T/ V99T	0.5 (0.1)	-11.4 (0.1)	-13.8	2.4	5.6 (0.1)	0.1 (0.1)

[a]All experiments were performed in 25 mM HEPES with 100 mM KCl at 25 °C, pH 7

[b]Values in parenthesis represent the error of the fit. A single experiment was performed for each protein.

Table 2: Thermodynamic values from pressure denaturation experiments

Protein	Conditions (M GdnHCl)	ΔG^{o}_{1bar} (kcal/mol)	ΔV (mL/mol)	P_{mid} (bar)
Δ+PHS	2.3	0.8 (0.1)	58 (4)	577 (310)
V23A	0.8	4.0 (0.2)	74 (4)	2262 (1076)
	1.0	3.3 (0.3)	74 (6)	1866 (1337)
	1.4	1.6 (0.2)	62 (7)	1080 (975)
L25A	1.0	3.8 (0.1)	89 (3)	1786 (575)
	1.2	3.2 (0.1)	96 (4)	1395 (575)
	1.4	2.1 (0.2)	87 (7)	1010 (771)
L36A	0.9	3.4 (0.2)	95 (4)	1497 (635)
	1.1	2.8 (0.1)	104 (4)	1126 (447)
	1.3	1.9 (0.2)	101 (7)	787 (560)
V66A	1.2	4.1 (0.1)	87 (3)	1972 (642)
	1.4	2.8 (0.2)	70 (4)	1674 (825)
	1.6	1.8 (0.1)	66 (3)	1141 (430)
I92A	0.8	3.4 (0.2)	110 (6)	1293 (753)
	1.0	2.8 (0.1)	113 (4)	1037 (395)
	1.2	2.1 (0.1)	130 (7)	676 (418)
I92N	0.2	3.2 (0.1)	84 (2)	1584 (357)
I92Q	0.4	3.3 (0.1)	85 (2)	1640 (367)
V23T/V99T	0.4	2.8 (0.1)	85 (2)	1374 (309)
V66T/V99T	0.9	2.1 (0.1)	53 (3)	1693 (707)
I92N/V66A	0.0	1.6 (0.1)	91 (4)	715 (305)
I92Q/V66A	0.0	3.0 (0.2)	91 (6)	1357 (868)
V23T/V66T/V99T	0.4	2.2 (0.1)	77 (4)	1196 (551)
V23T/V66A/V99T	0.2	2.3 (0.1)	82 (3)	1169 (395)
V23A/V66T/V99T	0.2	4.0 (0.1)	127 (3)	1323 (358)
V23T/L25A/V66T	0.3	3.4 (0.1)	111 (3)	1295 (375)
V23T/L25A/V99T	0.0	2.1 (0.1)	97 (4)	902 (372)
	0.1	2.0 (0.1)	120 (4)	703 (261)
L25A/V66T/V99T	0.3	3.4 (0.1)	106 (2)	1342 (271)
V23T/L36A/V66T	0.2	3.4 (0.1)	107 (3)	1336 (394)
	0.3	3.0 (0.1)	107 (4)	1169 (457)
V23T/L36A/V99T	0.0	2.7 (0.1)	87 (3)	1303 (427)
	0.1	2.7 (0.2)	96 (6)	1196 (747)
L36A/V66T/V99T	0.1	3.0 (0.1)	103 (4)	1209 (482)
	0.2	2.7 (0.1)	117 (4)	969 (363)
V23T/L25A/V66T/V99T	0.0	1.6 (0.1)	113 (7)	599 (397)
V23T/L36A/V66T/V99T	0.0	1.5 (0.1)	104 (7)	603 (417)
V23T/V66T/V74T/V99T	0.0	0.5 (0.2)	79 (8)	265 (250)

[a]All experiments were performed in 50 mM Tris at 20 °C, pH 7

Table 3: Volume calculations from crystal structures.

Protein	Change in van der Waals volume (mL/mol)	Hydration state of cavity	Probe radius (Å)	Calculated volume (mL/mol)		PDB ID
				McVol	CASTp	
Δ+PHS	0.00	crystal (empty)	1.2 1.4	40 11	84 39	3BDC*
V23A	-22.76	crystal (empty)	1.2 1.4	60 33	96 59	3PMF^
L25A	-33.73	crystal (empty)	1.2 1.4	58 7	108 30	3OSO^
L36A	-33.73	crystal (empty)	1.2 1.4	47 20	110 50	3NP8^
V66A	-22.76	crystal (empty)	1.2 1.4	92 30	123 63	3NQT^
V23S	-19.26	empty	1.2 1.4	70 39	130 77	4KHV
		crystal	1.2 1.4	70 39	130 77	
I92S	-30.70	crystal (empty)	1.2 1.4	57 24	122 74	4PMB
I92N	-16.86	crystal (empty)	1.2 1.4	53 23	128 58	4NDX
I92Q	-6.02	empty	1.2 1.4	28 --	119 29	4K2L
		crystal	1.2 1.4	23 --	102 16	
V23S/V66A	-42.14	empty	1.2 1.4	100 60	180 115	4ME5
		crystal	1.2 1.4	92 60	163 107	
V23Q/V66A	-17.46	empty	1.2 1.4	63 --	130 86	4NMZ
		crystal	1.2 1.4	51 --	111 78	
V23T/V66T	-14.45	empty	1.2 1.4	41 18	118 44	4RKL
		crystal	1.2 1.4	42 19	106 44	
V23T/V99T	-14.45	empty	1.2 1.4	47 19	121 58	4MIU
		crystal	1.2 1.4	49 19	108 58	
V66A/I92S	-53.58	crystal (empty)	1.2 1.4	94 59	184 140	4N9T
V66A/I92N	-39.73	empty	1.2 1.4	117 75	198 141	4NP5
		crystal	1.2 1.4	38 14	92 47	
V66A/I92Q	-28.90	empty	1.2 1.4	94 56	178 103	4NKL
		crystal	1.2 1.4	45 14	109 46	

Table 3 (continued)

Protein	Change in van der Waals volume (mL/mol)	Hydration state of cavity	Probe radius (Å)	Calculated volume (mL/mol)		PDB ID
				McVol	CASTp	
V66T/V99T	-14.45	crystal (empty)	1.2 1.4	26 9	78 32	4RKB
V23T/L25A /V99T	-48.76	empty	1.2 1.4	67 26	160 68	4N9P
		crystal	1.2 1.4	62 25	160 68	
V23T/L36A /V99T	-48.76	empty	1.2 1.4	47 28	122 58	4K5V
		crystal	1.2 1.4	14 --	74 21	
V23T/V66A /V99T	-37.32	empty	1.2 1.4	81 38	159 95	4KJN
		crystal	1.2 1.4	76 39	142 95	
L36A/V66T /V99T	-48.76	crystal (empty)	1.2 1.4	49 22	129 52	4KJO
V23T/L25A /V99T	-48.76	empty	1.2 1.4	67 26	160 68	4N9P
		crystal	1.2 1.4	62 25	160 68	
V23T/L36A /V99T	-48.76	empty	1.2 1.4	47 28	122 58	4K5V
		crystal	1.2 1.4	14 --	74 21	
V23T/V66A /V99T	-37.32	empty	1.2 1.4	81 38	159 95	4KJN
		crystal	1.2 1.4	76 39	142 95	
L36A/V66T /V99T	-48.76	crystal (empty)	1.2 1.4	49 22	129 52	4KJO

* Crystal structure from (68).

^ Crystal structure from (11).

Figure legends

Figure 1. Composite representation of conserved sites where internal water molecules are observed in variants with internal polar groups at positions 23, 66, 92 and 99. Water molecules are represented as blue spheres, and labeled according to the nomenclature found in (8). The Cα atom of positions 23, 66, 92 and 99 are highlighted in orange.

Figure 2. Crystal structures of the parent Δ+PHS variant of SNase and of variants with polar substitutions in core positions. Cavities are shown in red mesh calculated as a molecular surface in PyMOL (69) using a probe of 1.4 Å. PDB ID is in parenthesis. **(A)** Δ+PHS (3BDC), **(B)** V23S (4KHV), **(C)** I92S (4PMB), **(D)** I92N (4NDX), **(E)** I92Q (4K2L), **(F)** V23T/V66T (4RKL), **(G)** V23T/V99T (4MIU), **(H)** V66T/V99T (4RKB), **(I)** V23S/V66A (4ME5), **(J)** V23Q/V66A (4NMZ), **(K)** I92S/V66A (4N9T), **(L)** V66A/I92N (4NP5), **(M)** V66A/I92Q (4NKL), **(N)** V23T/L25A/V99T (4N9P), **(O)** V23T/L36A/V99T (4K5V), **(P)** V23T/V66A/V99T (4KJN), and **(Q)** L36A/V66T/V99T (4KJO).

Figure 3. Denaturation with GdnHCl monitored with Trp fluorescence for Δ+PHS SNase (bold) and for variants of Δ+PHS. All measurements were performed at pH 7 and 25 °C.

Figure 4. Normalized pressure unfolding profiles measured by following the center of spectral mass of Trp-140 fluorescence as a function of pressure. **(A)** L36A in 1.10M GdnHCl in filled circles (•), V23T/L36A/V66T in 0.25M GdnHCl in crosses (+), V23T/L36A/V99T in 0.05M GdnHCl in open squares (□), and L36A/V66T/V99T in

0.10M GdnHCl in filled triangles (▼). **(B)** V23T/V99T in 0.40M GdnHCl in filled circles (•), V23T/L25A/V99T in 0M GdnHCl in crosses (+), V23T/L36A/V99T in 0.05M GdnHCl in open squares (□), and V23T/V66A/V99T in 0.20M GdnHCl in filled triangles (▼). **(C)** V66T/V99T in 0.85M GdnHCl in filled circles (•), V23A/V66T/V99T in 0.20M GdnHCl in crosses (+), L25A/V66T/V99T in 0.30M GdnHCl in open squares (□), and L36A/V66T/V99T in 0.10M GdnHCl in filled triangles (▼). **(D)** I92N in 0.2M GdnHCl in filled circles (•), I92Q in 0.35M GdnHCl in crosses (+), V66A/I92N in 0M GdnHCl in open squares (□), and V66A/I92Q in 0M GdnHCl in filled triangles (▼). All measurements were performed at pH 7.

Figure 5. (A) Microenvironment of position 66, and **(B)** overlay of the C-α trace of Δ+PHS (white), V23T/V99T (red) and V66T/V99T (blue). Position 66, 17 and 67 are denoted. Water molecules are shown as small spheres and colored according to the structure they correspond to. Water sites 1, 3, 7 and 8 are denoted.

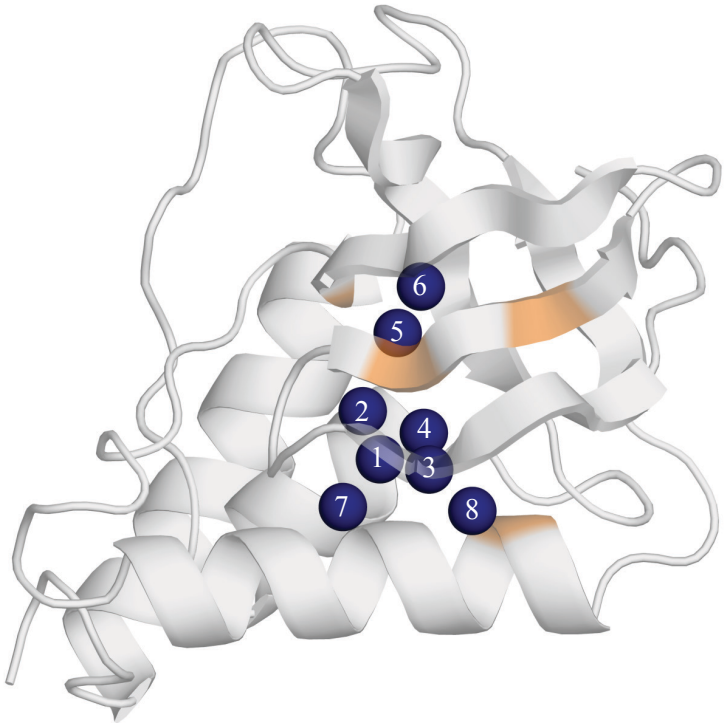


Figure 1

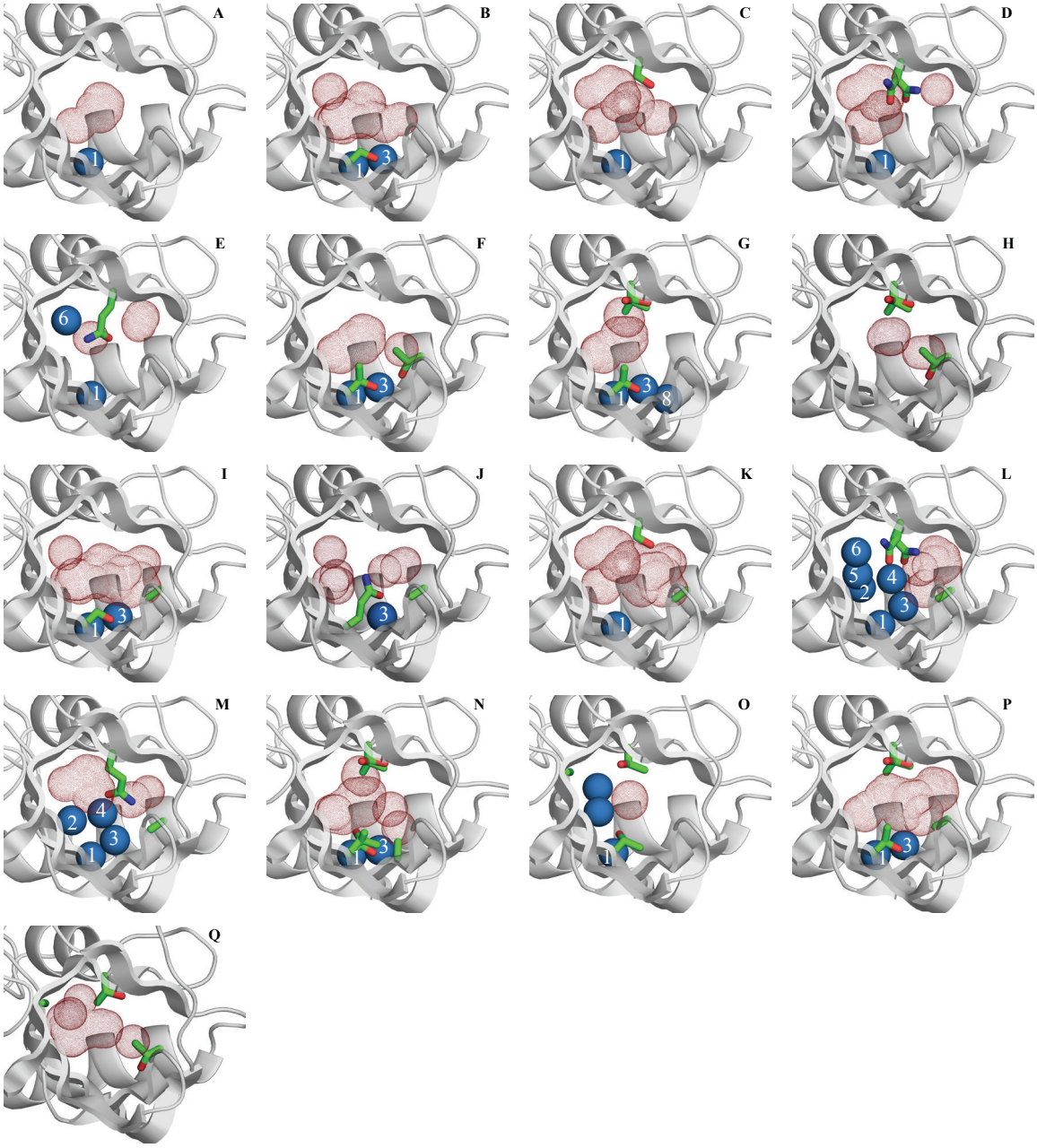


Figure 2

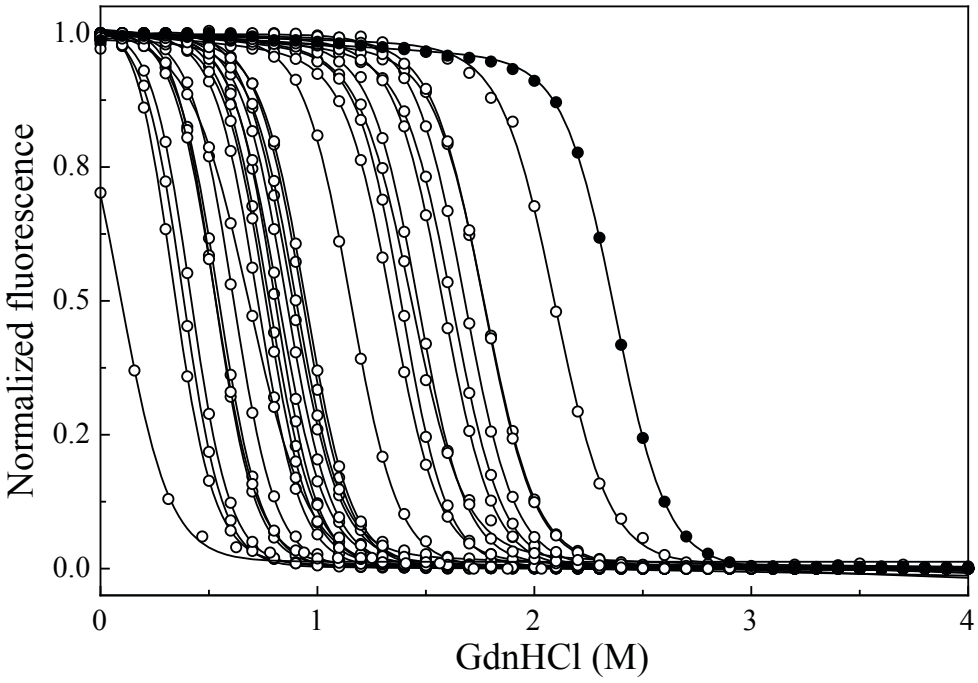


Figure 3

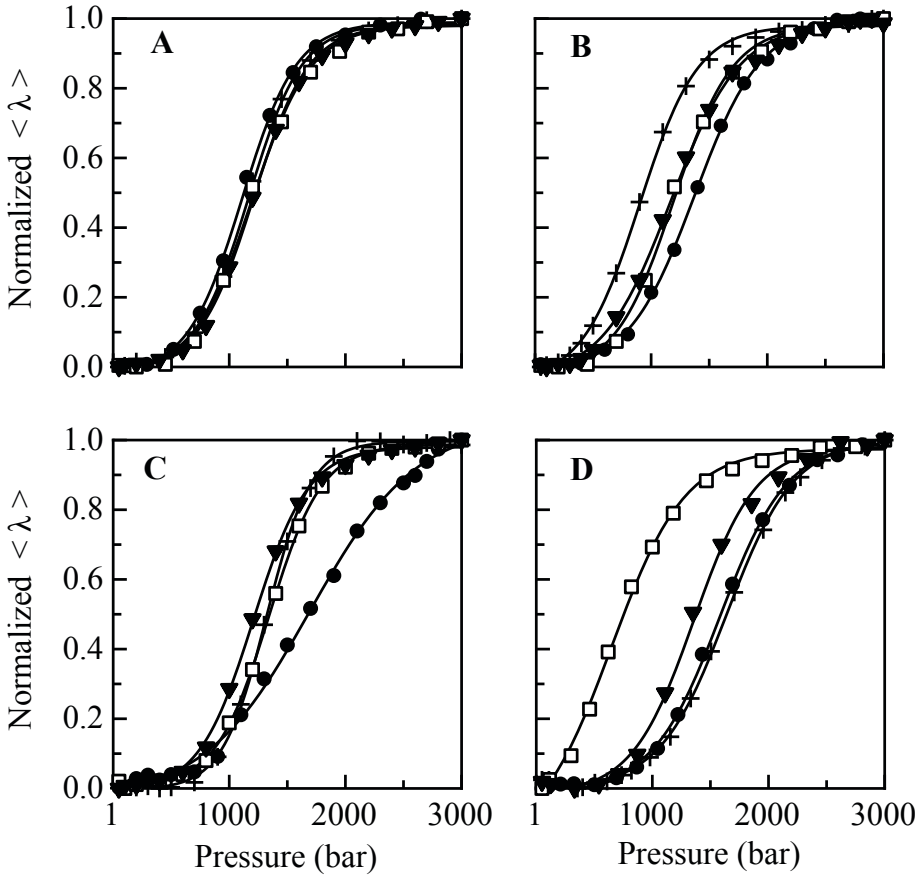

A
B
C
D
Normalized < λ >
Normalized < λ >
1.0
0.8
0.6
0.4
0.2
0.0
1
1000
2000
3000
Pressure (bar)
Pressure (bar)

Figure 4

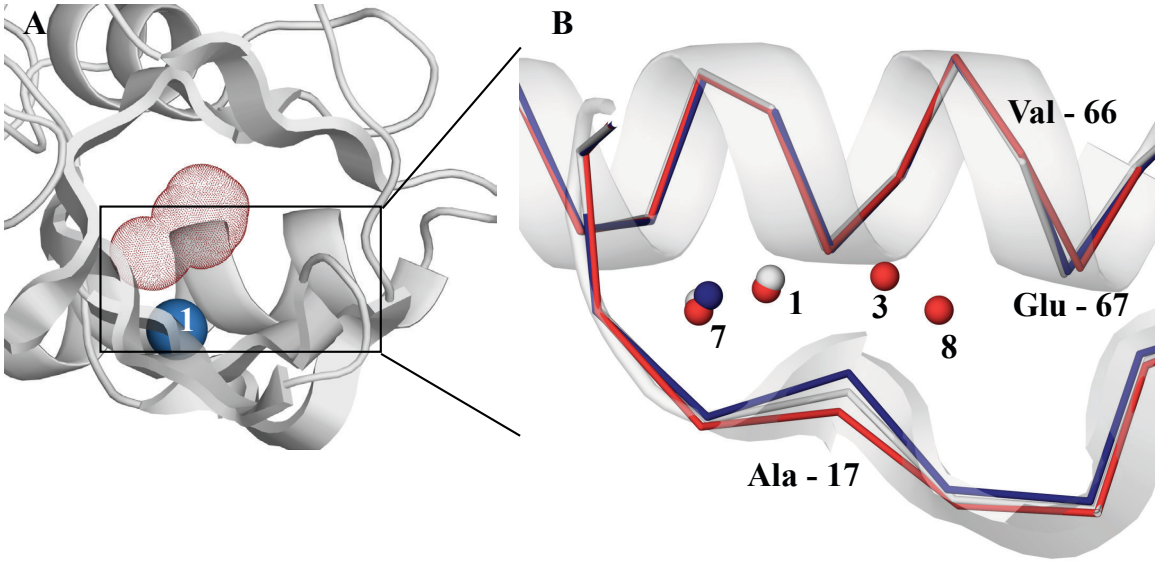


Figure 5

Supplemental Table 1: Crystallographic values

Crystallization conditions	Δ+PHS V23S	Δ+PHS I92S
Buffer	25mM potassium phosphate	25mM potassium phosphate
pH	9	8
Temperature (K)	277	277
Precipitant	26% (w/v) MPD	25% (w/v) MPD
Additives	pdTp, Calcium Chloride	pdTp, Calcium Chloride
Data collection		
Wavelength (Å)	1.1	1.54
Resolution (Å)	50.00-1.60 (1.63-1.60)	50.00-1.80 (1.83-1.80)
Unique reflections	18541 (907)	12978 (620)
Completeness	0.997 (0.974)	0.999 (1.000)
Redundancy	6.4 (5.3)	9.1 (5.7)
Average $I/\sigma(I)$	11.5 (7.8)	19.0 (5.1)
R_{merge}	0.033 (0.261)	0.035 (0.227)
Wilson B ($Å^2$)	33.3	27.2
Space group	$P2_1$	$P2_1$
Cell dimensions (Å ; °)	a = 31.15 ; α = 90.00 b = 60.40 ; β = 93.68 c = 37.91 ; γ = 90.00	a = 31.03 ; α = 90.00 b = 59.96 ; β = 93.63 c = 38.02 ; γ = 90.00
Refinement		
Resolution (Å)	32.06-1.60 (1.64-1.60)	37.94-1.80 (1.85-1.80)
No. of non-hydrogen atoms	1176	1154
No. of unique reflections	18521 (1294)	12950 (905)
No. of reflections in test set	947 (73)	630 (47)
R_{work}	0.169 (0.23)	0.196 (0.26)
R_{free}	0.207 (0.30)	0.227 (0.30)
RMSD from ideal geometry		
Bonds (Å)	0.016	0.018
RMS angles (°)	1.74	1.82
Average B-factors ($Å^2$)		
Protein	27.7	21.8
Solvent	32.2	23.0
Ion	26.3	19.7
Ramachandran plot		
Most favored (%)	101 (88.6)	99 (86.8)
Additionally allowed (%)	12 (10.5)	14 (12.2)
Disallowed (%)	1 (0.9)	1 (0.9)
No. of residues excluding Gly, Pro and termini	114	114
Total no. of residues	129	129
PDB accession code	4KHV	4PMB
RMSD (Å) from Δ+PHS	0.32	0.37
Main chain only	0.15	0.16

Supplemental Table 1 (continued)

Crystallization conditions	Δ+PHS I92N	Δ+PHS I92Q
Buffer	25mM potassium phosphate	25mM potassium phosphate
pH	9	8
Temperature (K)	277	277
Precipitant	20% (w/v) MPD	22% (w/v) MPD
Additives	pdTp, Calcium Chloride	pdTp, Calcium Chloride
Data collection		
Wavelength (Å)	1.1	1.1
Resolution (Å)	50.00-1.50 (1.53-1.50)	50.00-1.55 (1.58-1.55)
Unique reflections	22875 (1102)	20512 (999)
Completeness	0.994 (0.971)	0.997 (0.982)
Redundancy	6.1 (4.9)	6.1 (74.8)
Average $I/\sigma(I)$	13.9 (6.6)	18.5 (8.4)
R_{merge}	0.054 (0.223)	0.054 (0.191)
Wilson B ($Å^2$)	31.5	31.7
Space group	$P2_1$	$P2_1$
Cell dimensions (Å ; °)	a = 31.25 ; α = 90.00 b = 60.58 ; β = 93.15 c = 38.28 ; γ = 90.00	a = 31.05 ; α = 90.00 b = 60.27 ; β = 92.94 c = 38.34 ; γ = 90.00
Refinement		
Resolution (Å)	38.23-1.50 (1.53-1.50)	38.29-1.55 (1.59-1.55)
No. of non-hydrogen atoms	1276	1223
No. of unique reflections	22858 (1546)	20493 (1390)
No. of reflections in test set	1177 (83)	1051 (63)
R_{work}	0.184 (0.31)	0.172 (0.21)
R_{free}	0.216 (0.37)	0.214 (0.28)
RMSD from ideal geometry		
Bonds (Å)	0.020	0.019
RMS angles (°)	1.95	1.91
Average B-factors ($Å^2$)		
Protein	24.4	26.3
Solvent	32.5	32.1
Ion	25.4	23.9
Ramachandran plot		
Most favored (%)	100 (87.7)	98 (86.0)
Additionally allowed (%)	13 (11.4)	15 (13.2)
Disallowed (%)	1 (0.9)	1 (0.9)
No. of residues excluding Gly, Pro and termini	114	114
Total no. of residues	129	129
PDB accession code	4NDX	4K2L
RMSD (Å) from Δ+PHS	0.30	0.33
Main chain only	0.14	0.15

Supplemental Table 1 (continued)

Crystallization conditions	Δ+PHS V23T/V66T	Δ+PHS V23T/V99T
Buffer	25mM potassium phosphate	25mM potassium phosphate
pH	8	9
Temperature (K)	277	277
Precipitant	20% (w/v) MPD	20% (w/v) MPD
Additives	pdTp, Calcium Chloride	pdTp, Calcium Chloride
Data collection		
Wavelength (Å)	1.54	1.54
Resolution (Å)	50.00-1.66 (1.76-1.66)	50.00-1.67 (1.77-1.67)
Unique reflections	16877 (2701)	16555 (2612)
Completeness	1.000 (1.000)	0.992 (1.000)
Redundancy	20.4 (5.8)	11.7 (8.3)
Average $I/\sigma(I)$	44.5 (6.6)	46.5 (5.1)
R_{merge}	0.015 (0.118)	0.015 (0.184)
Wilson B ($Å^2$)	25.4	29.9
Space group	$P2_1$	$P2_1$
Cell dimensions (Å ; °)	a = 31.15 ; α = 90.00 b = 60.40 ; β = 94.48 c = 38.47 ; γ = 90.00	a = 31.20 ; α = 90.00 b = 60.55 ; β = 93.32 c = 38.20 ; γ = 90.00
Refinement		
Resolution (Å)	38.35-1.66 (1.70 -1.66)	38.14-1.67 (1.71-1.67)
No. of non-hydrogen atoms	1204	1243
No. of unique reflections	15965 (1203)	16528 (1192)
No. of reflections in test set	848 (59)	835 (59)
R_{work}	0.187 (0.22)	0.184 (0.23)
R_{free}	0.232 (0.29)	0.235 (0.30)
RMSD from ideal geometry		
Bonds (Å)	0.019	0.019
RMS angles (°)	1.98	1.91
Average B-factors ($Å^2$)		
Protein	19.0	23.6
Solvent	22.5	29.2
Ion	18.4	23.4
Ramachandran plot		
Most favored (%)	100 (87.7)	102 (89.5)
Additionally allowed (%)	13 (11.4)	11 (9.7)
Disallowed (%)	1 (0.9)	1 (0.9)
No. of residues excluding Gly, Pro and termini	114	114
Total no. of residues	129	129
PDB accession code	4RKL	4MIU
RMSD (Å) from Δ+PHS	0.30	0.33
Main chain only	0.15	0.15

Supplemental Table 1 (continued)

Crystallization conditions	Δ+PHS V66T/V99T	Δ+PHS V23S/V66A
Buffer	25mM potassium phosphate	25mM potassium phosphate
pH	9	6
Temperature (K)	277	277
Precipitant	30% (w/v) MPD	25% (w/v) MPD
Additives	pdTp, Calcium Chloride	pdTp, Calcium Chloride
Data collection		
Wavelength (Å)	1.54	1.54
Resolution (Å)	50.00-1.88 (1.97-1.88)	50.00-1.80 (1.89-1.80)
Unique reflections	11479 (1487)	13154 (1789)
Completeness	1.000 (1.000)	1.000 (1.000)
Redundancy	11.6 (6.4)	10.7 (4.9)
Average $I/\sigma(I)$	26.5 (4.2)	23.8 (3.7)
R_{merge}	0.026 (0.227)	0.033 (0.261)
Wilson B ($Å^2$)	32.6	26.7
Space group	$P2_1$	$P2_1$
Cell dimensions (Å ; °)	a = 30.98 ; α = 90.00 b = 60.54 ; β = 93.10 c = 37.95 ; γ = 90.00	a = 31.16 ; α = 90.00 b = 60.33 ; β = 93.61 c = 38.13 ; γ = 90.00
Refinement		
Resolution (Å)	37.89-1.88 (1.93-1.88)	38.05-1.80 (1.85-1.80)
No. of non-hydrogen atoms	1150	1197
No. of unique reflections	11462 (821)	13131 (929)
No. of reflections in test set	548 (34)	1306 (98)
R_{work}	0.185 (0.23)	0.164 (0.19)
R_{free}	0.225 (0.27)	0.222 (0.27)
RMSD from ideal geometry		
Bonds (Å)	0.019	0.019
RMS angles (°)	1.99	1.94
Average B-factors ($Å^2$)		
Protein	27.0	20.5
Solvent	29.4	24.4
Ion	27.3	22.6
Ramachandran plot		
Most favored (%)	99 (86.8)	101 (88.6)
Additionally allowed (%)	14 (12.3)	12 (11.4)
Generously allowed (%)	1 (0.9)	1 (0.9)
Disallowed (%)		
No. of residues excluding	114	114
Gly, Pro and termini	129	129
Total no. of residues		
PDB accession code	4RKB	4ME5
RMSD (Å) from Δ+PHS	0.35	0.34
Main chain only	0.17	0.15

Supplemental Table 1 (continued)

Crystallization conditions	Δ+PHS V23Q/V66A	Δ+PHS V66A/I92S
Buffer	25mM potassium phosphate	25mM potassium phosphate
pH	8	8
Temperature (K)	277	277
Precipitant	20% (w/v) MPD	20% (w/v) MPD
Additives	pdTp, Calcium Chloride	pdTp, Calcium Chloride
Data collection		
Wavelength (Å)	1.1	1.1
Resolution (Å)	50.00-1.70 (1.73-1.70)	50.00-1.60 (1.63-1.60)
Unique reflections	15361 (773)	18603 (912)
Completeness	0.977 (0.981)	0.993 (0.996)
Redundancy	6.2 (6.3)	5.4 (5.3)
Average $I/\sigma(I)$	20.5 (16.4)	9.2 (17.7)
R_{merge}	0.049 (0.132)	0.086 (0.267)
Wilson B ($Å^2$)	29.8	31.1
Space group	$P2_1$	$P2_1$
Cell dimensions (Å ; °)	a = 31.07 ; α = 90.00 b = 60.42 ; β = 94.43 c = 38.33 ; γ = 90.00	a = 31.15 ; α = 90.00 b = 60.35 ; β = 93.72 c = 38.24 ; γ = 90.00
Refinement		
Resolution (Å)	38.22-1.70 (1.79-1.74)	38.16-1.60 (1.64-1.60)
No. of non-hydrogen atoms	1252	1226
No. of unique reflections	15256 (1094)	18585 (1348)
No. of reflections in test set	770 (59)	950 (64)
R_{work}	0.173 (0.21)	0.189 (0.15)
R_{free}	0.212 (0.31)	0.223 (0.28)
RMSD from ideal geometry		
Bonds (Å)	0.019	0.019
RMS angles (°)	1.98	1.99
Average B-factors ($Å^2$)		
Protein	23.5	22.4
Solvent	30.5	29.4
Ion	24.8	24.4
Ramachandran plot		
Most favored (%)	101 (88.6)	98 (86.0)
Additionally allowed (%)	12 (10.5)	15 (13.2)
Disallowed (%)	1 (0.9)	1 (0.9)
No. of residues excluding Gly, Pro and termini	114	114
Total no. of residues	129	129
PDB accession code	4NMZ	4N9T
RMSD (Å) from Δ+PHS	0.34	0.33
Main chain only	0.16	0.15

Supplemental Table 1 (continued)

Crystallization conditions	Δ+PHS V66A/I92N	Δ+PHS V66A/I92Q
Buffer	25mM potassium phosphate	25mM potassium phosphate
pH	7	6
Temperature (K)	277	277
Precipitant	20% (w/v) MPD	25% (w/v) MPD
Additives	pdTp, Calcium Chloride	pdTp, Calcium Chloride
Data collection		
Wavelength (Å)	1.1	1.1
Resolution (Å)	50.00-1.50 (1.53-1.50)	50.00-1.60 (1.63-1.60)
Unique reflections	22825 (1123)	18886 (926)
Completeness	0.997 (0.989)	0.999 (0.997)
Redundancy	9.2 (9.2)	6.3 (5.6)
Average $I/\sigma(I)$	16.4 (16.3)	17.1 (12.1)
R_{merge}	0.062 (0.155)	0.051 (0.149)
Wilson B ($Å^2$)	27.2	30.0
Space group	$P2_1$	$P2_1$
Cell dimensions (Å ; °)	a = 31.12 ; α = 90.00 b = 60.45 ; β = 93.79 c = 38.54 ; γ = 90.00	a = 31.19 ; α = 90.00 b = 60.48 ; β = 93.96 c = 38.17 ; γ = 90.00
Refinement		
Resolution (Å)	38.45-1.50 (1.54-1.50)	38.07-1.60 (1.64-1.60)
No. of non-hydrogen atoms	1246	1214
No. of unique reflections	22810 (1582)	18708 (1361)
No. of reflections in test set	1171 (95)	965 (81)
R_{work}	0.181 (0.22)	0.174 (0.21)
R_{free}	0.214 (0.31)	0.219 (0.28)
RMSD from ideal geometry		
Bonds (Å)	0.018	0.020
RMS angles (°)	1.95	1.98
Average B-factors ($Å^2$)		
Protein	21.0	22.7
Solvent	27.4	29.1
Ion	20.3	22.4
Ramachandran plot		
Most favored (%)	100 (87.7)	98 (86.0)
Additionally allowed (%)	13 (11.4)	15 (13.2)
Disallowed (%)	1 (0.9)	1 (0.9)
No. of residues excluding Gly, Pro and termini	114	114
Total no. of residues	129	129
PDB accession code	4NP5	4NKL
RMSD (Å) from Δ+PHS	0.31	0.31
Main chain only	0.16	0.17

Supplemental Table 1 (continued)

Crystallization conditions	Δ+PHS V23T/L25A/V99T	Δ+PHS V23T/L36A/V99T
Buffer	25mM potassium phosphate	25mM potassium phosphate
pH	8	8
Temperature (K)	277	277
Precipitant	25% (w/v) MPD	20% (w/v) MPD
Additives	pdTp, Calcium Chloride	pdTp, Calcium Chloride
Data collection		
Wavelength (Å)	1.1	1.1
Resolution (Å)	50.00-1.40 (1.42-1.40)	50.00-1.75 (1.78-1.75)
Unique reflections	26893 (1332)	14288 (686)
Completeness	0.962 (0.932)	0.992 (0.979)
Redundancy	6.3 (5.3)	6.3 (6.4)
Average $I/\sigma(I)$	18.8 (12.3)	14 (12.1)
R_{merge}	0.050 (0.153)	0.063 (0.175)
Wilson B ($Å^2$)	21.9	33.2
Space group	$P2_1$	$P2_1$
Cell dimensions (Å ; °)	a = 31.23 ; α = 90.00 b = 60.13 ; β = 94.72 c = 38.34 ; γ = 90.00	a = 31.15 ; α = 90.00 b = 60.45 ; β = 93.11 c = 38.37 ; γ = 90.00
Refinement		
Resolution (Å)	38.21-1.40 (1.43-1.40)	38.31-1.75 (1.79-1.75)
No. of non-hydrogen atoms	1250	1196
No. of unique reflections	26872 (1848)	14279 (960)
No. of reflections in test set	1351 (67)	720 (45)
R_{work}	0.177 (0.23)	0.168 (0.15)
R_{free}	0.196 (0.25)	0.221 (0.28)
RMSD from ideal geometry		
Bonds (Å)	0.020	0.020
RMS angles (°)	1.96	1.93
Average B-factors ($Å^2$)		
Protein	17.8	27.6
Solvent	26.1	30.9
Ion	17.1	30.4
Ramachandran plot		
Most favored (%)	100 (87.7)	100 (87.7)
Additionally allowed (%)	13 (11.4)	13 (11.4)
Disallowed (%)	1 (0.9)	1 (0.9)
No. of residues excluding Gly, Pro and termini	114	114
Total no. of residues	129	129
PDB accession code	4N9P	4K5V
RMSD (Å) from Δ+PHS	0.39	0.33
Main chain only	0.18	0.17

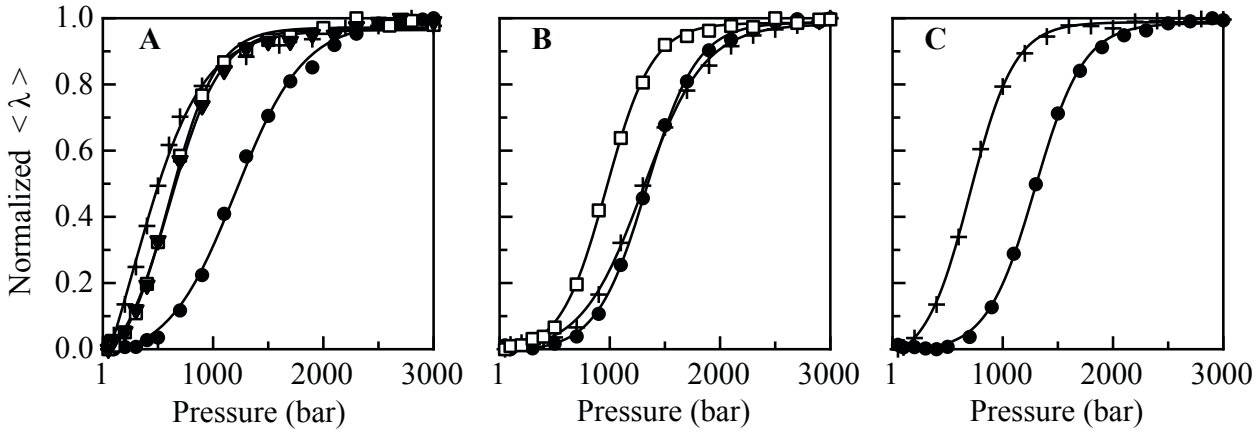
Supplemental Table 1 (continued)

Crystallization conditions	Δ+PHS V23T/V66A/V99T	Δ+PHS L36A/V66T/V99T
Buffer	25mM potassium phosphate	25mM potassium phosphate
pH	8	6
Temperature (K)	277	277
Precipitant	20% (w/v) MPD	20% (w/v) MPD
Additives	pdTp, Calcium Chloride	pdTp, Calcium Chloride
Data collection		
Wavelength (Å)	0.9790	1.54
Resolution (Å)	50.00-1.55 (1.58-1.55)	50.00-1.90 (1.94-1.90)
Unique reflections	20536 (994)	11300 (529)
Completeness	0.992 (1.000)	1.000 (1.000)
Redundancy	6.3 (6.5)	12.46 (8.76)
Average $I/\sigma(I)$	24 (16.4)	24.06 (3.71)
R_{merge}	0.044 (0.121)	0.028 (0.256)
Wilson B ($Å^2$)	29.1	34.2
Space group	$P2_1$	$P2_1$
Cell dimensions (Å ; °)	a = 31.08 ; α = 90.00 b = 60.04 ; β = 94.88 c = 38.48 ; γ = 90.00	a = 31.04 ; α = 90.00 b = 60.40 ; β = 94.72 c = 38.64 ; γ = 90.00
Refinement		
Resolution (Å)	38.34-1.55 (1.59-1.55)	38.51-1.90 (1.95-1.90)
No. of non-hydrogen atoms	1264	1143
No. of unique reflections	20212 (1334)	11275 (814)
No. of reflections in test set	1036 (82)	538 (41)
R_{work}	0.178 (0.20)	0.192 (0.26)
R_{free}	0.228 (0.22)	0.231 (0.36)
RMSD from ideal geometry		
Bonds (Å)	0.019	0.018
RMS angles (°)	1.98	1.91
Average B-factors ($Å^2$)		
Protein	22.9	37.4
Solvent	30.3	34.1
Ion	24.1	37.6
Ramachandran plot		
Most favored (%)	100 (87.7)	98 (86.0)
Additionally allowed (%)	13 (11.4)	15 (13.2)
Generously allowed (%)	1 (0.9)	1 (0.9)
Disallowed (%)		
No. of residues excluding	114	114
Gly, Pro and termini	129	129
Total no. of residues		
PDB accession code	4KJN	4KJO
RMSD (Å) from Δ+PHS	0.37	0.33
Main chain only	0.20	0.18

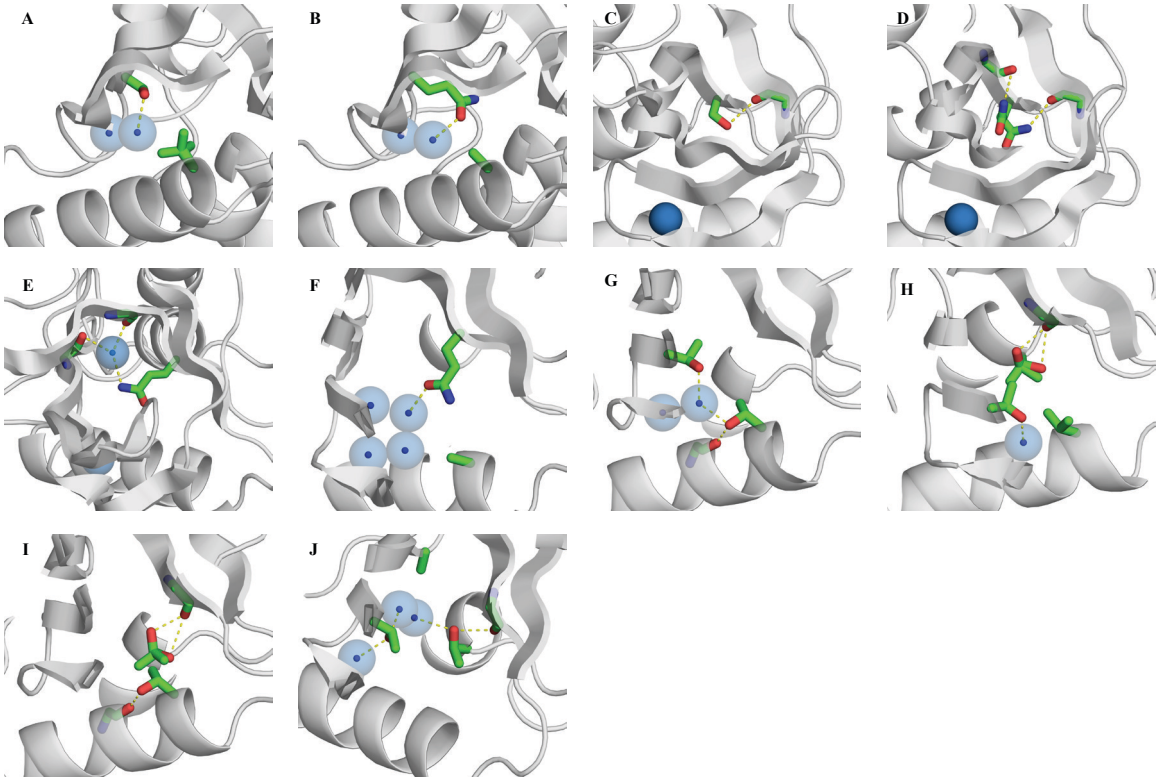
Supplemental Figure 1. Pressure unfolding profiles monitored by Trp fluorescence. **(A)** V23T/V66T/V99T in 0.4 M GdnHCl in filled circles (•), V23T/V66T/V74T/V99T in 0 M GdnHCl in crosses (+), V23T/L25A/V66T/V99T in 0 M GdnHCl in open squares (□), and V23T/L36A/V66T/V99T in 0 M GdnHCl in filled triangles (▼). **(B)** V23T/L36A/V66T in 0.2 M GdnHCl in filled circles (•), V23T/L36A/V99T in 0 M GdnHCl in crosses (+), L36A/V66T/V99T in 0.2 M GdnHCl in open squares (□). **(C)** V23T/L25A/V66T in 0.3 M GdnHCl in filled circles (•), V23T/L25A/V99T in 0.1 M GdnHCl in crosses (+). All measurements were performed at pH 7.

Supplemental Figure 2. Microenvironments of internal polar residues in SNase. **(A)** Structure of variant V23S showing a hydrogen bond between Ser-23 and a water molecule in site 3, as well as disorder in the side chain of Val-66. **(B)** Structure of variant V23Q/V66A showing a hydrogen bond between Gln-23 and a water molecule in site 3, as well as the Ala-66 side chain. **(C)** Structure of variant I92S showing a hydrogen bond between Ser-92 and the backbone carbonyl oxygen of Glu-73. **(D)** Structure of variant I92N showing a hydrogen bond between the main conformation of Asn-92 and the backbone carbonyl oxygen of Glu-73, a non-optimal polar contact of 3.3 Å between the secondary conformation of Asn-92 and the backbone carbonyl oxygen of Tyr-91. **(E)** Structure of variant I92Q showing a hydrogen bond between Gln-92 and a water molecule at site 6, as well as the hydrogen bonds between that water molecule and the backbone carbonyl oxygen of residues Arg-35 and Ala-90. **(F)** Structure of variant V66A/I92N showing a hydrogen bond between Gln-92 and a water molecule at site 4, which in turn is hydrogen bonded to water at sites 1, 2 and 3. The side chain of Ala-66 is

also shown. **(G)** Structure of variant V23T/V66T showing a hydrogen bond between Thr-23 and a water molecule at site 3. Thr-66 forms a hydrogen bond to the same water molecule and to the backbone carbonyl oxygen of Thr-62. **(H)** Structure of variant V23T/V99T showing a hydrogen bond between Thr-23 and a water molecule at site 3. Thr-99 is shown in two conformations, both forming a hydrogen bond with the backbone carbonyl oxygen of Ile-92. Val-66 is shown as well. **(I)** Structure of variant V66T/V99T showing a hydrogen bond between V66T and the backbone carbonyl oxygen of Thr-62, as well as a hydrogen bond between the two conformations of Thr-99 and the backbone carbonyl oxygen of Ile-92. **(J)** Structure of variant V23T/L36A/V99T showing the hydrogen bond network formed by Thr-23, a water molecule at site 1, two previously not observed internal water molecules, and Thr-99, which also forms a hydrogen bond with the backbone carbonyl oxygen of Ile-92.



Supplemental Figure 1



Supplemental Figure 2

CHAPTER 5:

MATERIALS AND METHODS

Protein engineering. Experimental studies were performed using the WT and the highly stable Δ+PHS variant of SNase (1, 2) as the parent proteins, and variants thereof containing substitutions to Ala, to bulky hydrophobic side chains, or to polar side chains. The Δ+PHS background protein has the mutations of P117G, H124L, S128A, G50F, and V51N and the deletion of residues 44-49. The variant plasmids were prepared using site-directed mutagenesis with the QuikChange™ method (Integrated DNA Technologies, Coralville, IA) on a Novagen Pet24a+ (Merck Biosciences) vector as described previously (1, 2).

Protein expression and purification. Protein purification was performed as described previously (3). Unlabeled protein was produced by growing E.coli in lysogeny broth (LB) at 37 °C(4). Protein expression was controlled by phage T7 promoter system that is inducible with IPTG(5–7). Protein concentration was determined by measuring the absorbance of a protein sample through a 1 cm path at 280 nm using Cary 3C UV-visible spectrophotometer (Plainview, NY). The extinction coefficient that was used in this determination was 0.93 O.D. mL/mg (8). Protein purity was determined by SDS-PAGE analysis (9).

Crystallization and X-ray diffraction data collection. SNase variants were crystallized using the hanging drop vapor diffusion method (10). The protein was pre-incubated with calcium chloride and pdTp in a 1:3:2 molar equivalent ratio prior to mixing with the reservoir solution. These additives were not present in any of the thermodynamic equilibrium experiments. Despite involving the surface of the protein, it is possible that binding the inhibitor pdTp or the calcium might suppress structural changes that lead to

formation of cavities. However, crystal structures of SNase in the absence of inhibitor and calcium are almost indistinguishable and some examples of this can be found in PDB ID 3EJI (11) and 1SNP (12), and no evidence of any significant structural change is observed that might lead to changes in the volume of internal cavities.

In each crystal tray, the reservoir solution contained 25 mM potassium phosphate and 20% w/v 2-methyl-2,4-pentanediol (MPD) (Sigma-Aldrich Corp., St. Louis, MO) in the pH 6-9 range. The proteins were mixed in a 1:1 ratio with reservoir solutions prior to suspension over the reservoir solution and incubation at 277 K. Crystals appeared after 2 weeks or more. Single crystals were harvested using a nylon loop mounted on a copper base (CryoLoops™ and CrystalCap Copper Magnetic™ from Hampton Research, Aliso Viejo, CA) and flash-cooled in liquid nitrogen. Diffraction data were collected at 100K from a single crystal of each variant using the X25 beamline at Brookhaven National Laboratory (BNL) at wavelength 1.1000Å for Δ+PHS variants L25A/I92A, L36A/I92A, V74A/I92A, I92A/L103A, I92A/L125A, V23S, I92N, I92Q, V23Q/V66A, V66A/I92S, V66A/I92N, V66A/I92Q, V23T/L25A/V99T and V23T/L36A/V99T, and WT variants V23I/V66I/V74I/V99I, V23M/L36F and V23M/T62F; at wavelength 1.000Å for WT variant V23M/L25F/T62F; and at wavelength 0.979Å for Δ+PHS variant V23A/I92A and WT variants V23T/V66A/V99T, V23M, V23L/L25V/V66I/I72V and V23I/V66I/I72V/I92V. The data were indexed, integrated, scaled and merged using the HKL2000 software package (13) to yield a set of unique reflections. Diffraction data for Δ+PHS variant V66A/I92A, I92S, V23S/V66A, V23T/V66T, V23T/V99T, V66T/V99T, and L36A/V66T/V99T were collected on a Kappa ApexII diffractometer at wavelength 1.54Å, outfitted with a sealed copper tube, multilayer optics, and a CCD detector (Bruker

AXS, Madison, WI). Reflections were indexed, integrated and merged using the manufacturer's software ApexII and XPREP.

Crystallographic structure determination and analysis. Initial phases for each structure were obtained by maximum likelihood-based molecular replacement method with Phaser software (14) within the CCP4i (15) suite version 6.2.0, using the structure of Δ+PHS (PDB ID: 3BDC) as the search model for variants thereof, and the structure of WT (PDB ID: 1SNC) as the search model for variants thereof. Prior to molecular replacement, the search model coordinates were modified by truncating the substituted amino acid for the appropriate variant to Ala, removing all water molecules, and setting all B-factors to 20.00 $Å^2$. Residues H8, T13, Q30, K64, V74, T82, Y113, V114 and Y115 were also truncated to Ala in the search model. Iterative model building and refinement were performed using COOT (16) and Refmac5 (17) to yield the final models. Water molecules were added in the protein model during model building in regions where spherical electron density in 2Fo-Fc and Fo-Fc maps within 3.5 Å of a hydrogen bonding partner were observed. Structure factors, geometry and Ramachandran statistics were evaluated using the SFCHECK (18) and PROCHECK (18) programs, as well as the COOT validation suite and the MolProbity server (19). For structures of Δ+PHS variants V23A/I92A, L25A/I92A, L36A/I92A, V66A/I92A, V74A/I92A, I92A/L103A, I92A/L125A, V23S, I92S, I92Q, V23S/V66A, V23T/V66T, V23T/L36A/V99T and L36A/V66T/V99T, and for structures of WT variants V23I/V66I/I72V/I92V, V23I/V66I/V74I/V99I, V23L/L25V/V66I/I72V, V23M and V23M/L25F/T62F the TLSMD server (20) was used for the final combined TLS and restrained refinement cycle. Each structure reports Ramachandran statistics with upwards of 99% of the non-

glycine, non-proline and non-end residues in the allowed regions. Calculations for RMS deviation of protein atoms of the variant structure and the parent structure (either Δ+PHS or WT) were performed using the LSQKAB program (21). For side chains adopting multiple conformations, only the dominant conformer was included. In addition, for RMS calculations using the WT structure, amino acids 43-50 were excluded from the calculation since they belong to a loop that is highly disordered and usually left unmodeled.

Molecular dynamics. Ten nanosecond simulations were run on variants L36A/I92A and L25A/L36A/I92A to evaluate the range of volumes the cavities experience in solution. The 10ns trajectory was run using the PDB deposited structures, completed with termini generated using COOT (16), and run on Gromacs 4.6.1 (22) using the Amber99SB forcefield (23). One frame every 0.1 ns were chosen for a total of 100 frames to represent the trajectory and used to calculate the volume and standard deviation of the cavities detected using the McVol algorithm (24) and a probe radius of 1.4 Å or as specified.

Stability Measurements by Gdn unfolding. Stability measurements were performed with guanidine hydrochloride (Invitrogen, Carlsbad, CA) titrations using an Aviv automated titration differential spectrophotometer 105 (Lakewood, NJ) to monitor the intrinsic fluorescence of Trp-140 as a proxy for the foldedness of the protein. The measurements were performed at pH 7 with the excitation wavelength set to 296 nm and the emission wavelength to 326 nm, and the bandwidth for both set to 6.4 nm. Above pH 9, to avoid photobleaching, the excitation bandwidth was set to 1.0 nm. The protein sample included 25 mM HEPES, 100 mM KCl and approximately 3 nM of protein. Titrant consisted of 6M GdnHCl solution, 100 mM KCl, and 25 mM of HEPES. A 2 mL

protein sample was titrated with GdnHCl using Microlab 500 Pump (Hamilton) at 0.1 M increments. Final titrant concentration was verified by refractometry (25) using 334610 Thermo Electron refractometer. The final signal was adjusted for dilution and buffer effects. All experiments were carried out at room temperature (25 °C). For variants L25A/L36A/I92A, V23T/V66T/V74T/V99T and WT V23M/L25F/T62F, the stabilities were too low to determine accurately and refolding experiments were performed using ammonium sulfate ($AmSO_4$) to obtain native baseline points with which to fit the unfolding data, following the example in (26). The experiments were performed using the same spectrophotometer under the same settings as in the GdnHCl unfolding experiment of the corresponding variant and performed immediately following the end of the GdnHCl titrations.

A 2-state model was used to analyze the data, as previously employed (27), with y_N and y_U corresponding to signals from the native and unfolded states, respectively, and a linear dependence of the baseline signal as a function of denaturant concentration, such that for baseline i:

$$y_i = m_i x + b_i$$

With the fraction of protein in the native and unfolded states denoted by f_N and f_U and $f_N + f_U = 1$, the overall signal is composed then by:

$$y_{obs} = f_N y_N + f_U y_U$$

and

$$f_N = \frac{y_U - y_{obs}}{y_N - y_U} \; ; \; f_U = \frac{y_N - y_{obs}}{y_N - y_U}.$$

Relating the signal to the thermodynamic free energy of unfolding, given by:

$$\Delta G_{N \leftrightharpoons U} = -RT \ln K$$

where the equilibrium constant K is given by the ratio of unfolded to native protein and can be related to the fluorescence signal:

$$K = \frac{U}{N} = \frac{f_U}{f_N} = \frac{y_N - y_{obs}}{y_{obs} - y_U}.$$

Rewriting the observable in terms of the free energy, one gets:

$$y_{obs} = \frac{y_N + y_D \times e^{-\Delta G_{N \leftrightharpoons U}/RT}}{1 + e^{-\Delta G_{N \leftrightharpoons U}/RT}}.$$

A linear dependence of protein stability with denaturant concentration (28) was assumed, such that:

$$\Delta G_{N \leftrightharpoons U} = \Delta G^{\circ}_{N \leftrightharpoons U} - m_{N \leftrightharpoons U}[Denaturant]$$

where $\Delta G^{\circ}_{N \leftrightharpoons U}$ is the stability of the protein when $[Denaturant] = 0$ M.

The use of amino acid substitutions in proteins can lead to increased population of intermediate states and a departure from 2-state behavior (29). Studies of SNase and variants of it with an internal substitution to an ionizable residue showed that, for a small subset, the substitution induced conformational reorganization (30–32). The extent of reorganization appeared to be dependent on global stability and pH. The variants studied in this dissertation report stabilities ranging from 12 kcal/mol to less than 1 kcal/mol and involve substitutions that generate a very large cavity in the protein interior or induce extensive hydration of the protein core. It is possible that these substitutions lead to conformational reorganization and a significant deviation from 2-state behavior that cannot be detected by fluorescence but could become apparent with other techniques. The signal of Trp-140 in SNase is an excellent reporter of the global conformational state and recapitulates the unfolding transition obtained by other methods such as circular dichroism (33). Measurements by NMR spectroscopy presented in this dissertation begin

to reveal the heterogeneity of the response to pressure of SNase and its variants and shed light onto the local conformational rearrangements induced by the substitutions. Site-specific side chain and backbone information obtained by NMR spectroscopy already proves to be a powerful tool to explore the local effects of cavity-generating substitutions in T4 lysozyme (34, 35). The variants used in the systematic study presented in this dissertation are excellent candidates for a comprehensive exploration of the effect of cavities and internal polar groups on the folding energy landscape of SNase.

Pressure unfolding monitored by fluorescence. Experiments were carried out using an ISS steady-state fluorometer (Champaign, IL), with protein concentrations at 100 μM, except for variants V74A/I92A, I92A/L103A and L25A/L36A/I92A, 700 μM, 175 μM and 200 μM, respectively, and in 50 mM Tris buffer pH 7. A water bath stabilized the cell holder temperature at 293K. For a curve, each point represents the center of mass of a spectrum of Trp-140 from 320 to 450 nm, excited at 290 nm, or the average of the center of mass of several spectra to reduce uncertainty inherent to the apparatus. Measurements were made at equilibrium at pressures ranging from 50 to 3000 bar. Equilibrium was evaluated by monitoring the kinetic relaxation curves of the intensity at 340 nm (36). The center of mass of a spectrum of range *j* and at a given pressure *p* was obtained using an in-house program that calculates

$$< \lambda >_p = \frac{\sum_j F_j \lambda_j}{\sum_j F_j}$$

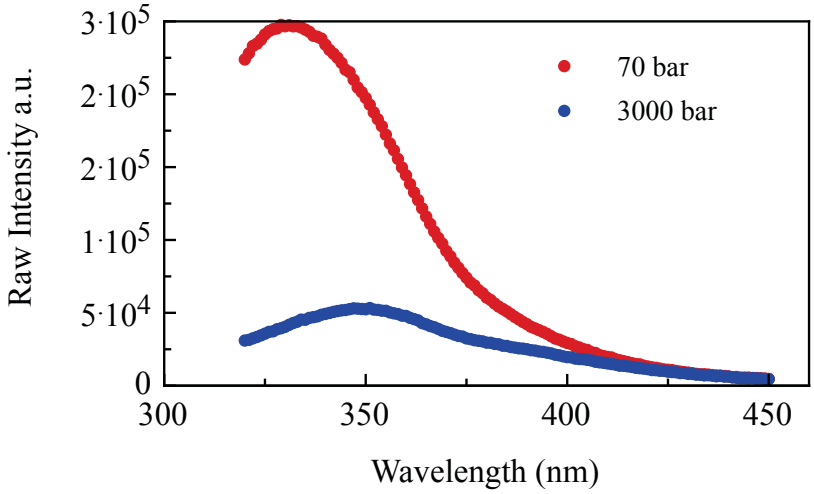


Figure 1. Fluorescence spectrum of SNase. Raw fluorescence intensity recorded as a function of wavelength at two different pressures: 70 bar (red) and 3000 bar (blue). The wavelength-weighted integral of this spectrum yields the center of spectral mass <λ>, which shifts to higher values as the fluorescence peak shifts to higher wavelengths. At high pressures, where the protein is unfolded, the signal is greatly reduced and the noise in the measurement increases.

Curves were fitted to a two-state unfolding model to obtain ΔG^o and ΔV^o using pro Fit v6.2.9 (QuantumSoft, Uetikon am See, Switzerland), according to

$$y_{obs} = \frac{y_N + y_D \times e^{-\Delta G_{N \leftrightarrow U}/RT}}{1 + e^{-\Delta G_{N \leftrightarrow U}/RT}}.$$

where

$$\Delta G_{N \leftrightharpoons U} = \Delta G^\circ{}_{N \leftrightharpoons U} + p\Delta V^\circ{}_{N \leftrightharpoons U}$$

and $\Delta G^\circ{}_{N \leftrightarrow U}$ corresponds to the stability of the protein at 1 bar. In this case, y_N and y_D are not linear functions of pressure but are instead fixed constants. This is because when data are collected in the pre- and post-denaturation regime, the signal of the baselines is observed to flatten out.

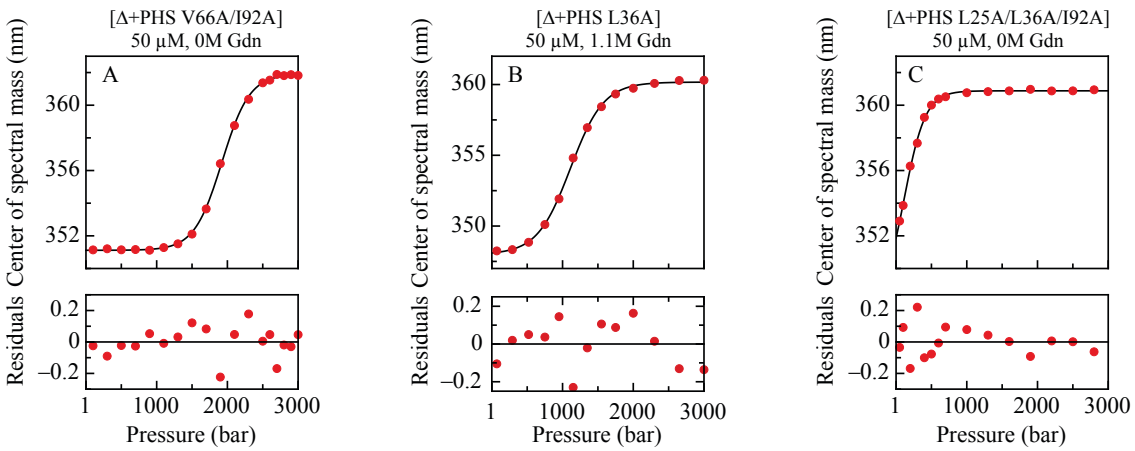
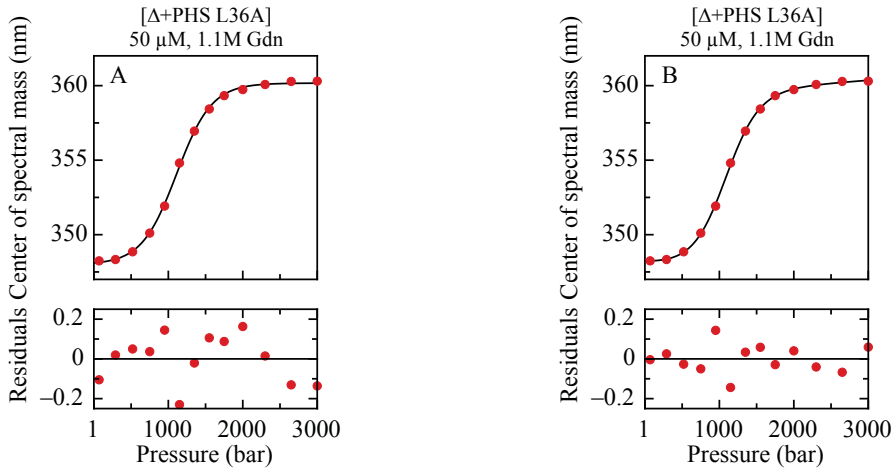


Figure 2. Folded and unfolded baselines as a function of pressure. Typical unfolding profiles of SNase are shown with the midpoint of the transition shifted relative to the accessible pressure range. The native (A) and unfolded (C) baselines do not show a dependence on pressure. When the midpoint of the transition coincides with the middle of the accessible pressure range, enough points of both baselines are captured.

It is also important to note the limited number of data points in the limited range of accessible pressures with respect to the number of parameters required for fitting. In an optimal dataset, the data points will reflect the pre-transitional, transitional, and post-transitional regions (Fig. 2A, B and C) of the unfolding behavior. However, the accessible pressure range limits the information to have limited information on baselines (Fig. 2B). A linear function for the baselines results in fitting errors that are comparable in magnitude to the baseline value itself. Still, for high quality data, the ΔV values recovered are within fitting error (Fig. 3A and 3B).

There exists in the data a small but reproducible deviation from a simple linear 2-state model that can be seen in the inflection region for the unfolded state (Fig. 3A). It is observed in multiple datasets and is sometimes less pronounced. This could reflect a true deviation from a 2-state transition or the detection of some compression effect. In order to address this, the fitting model was changed to both a compression model and a 3-state model (Fig. 4A).

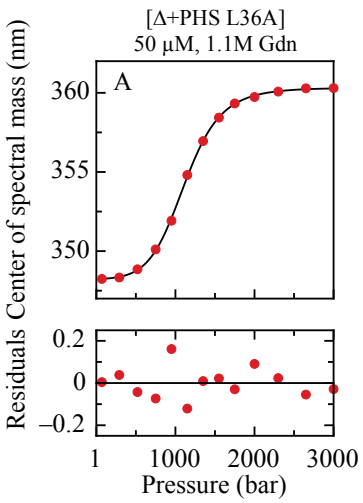


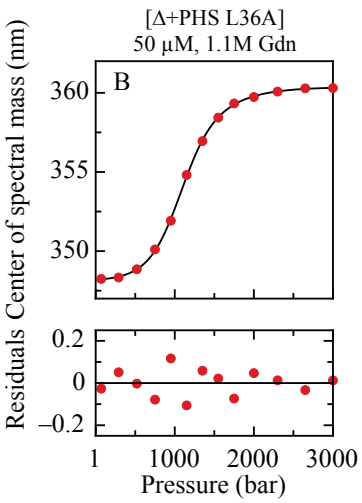
Fitted parameters:			Standard deviations:		
ΔV	=	103.7546	ΔΔV	=	3.5693
ΔG	=	-2756.8045	ΔΔG	=	102.2824
F	=	347.9960	ΔF	=	0.1123
U	=	360.1700	ΔU	=	7.2699e-2

Fitted parameters:			Standard deviations:		
ΔV	=	109.0030	ΔΔV	=	5.6886
ΔG	=	-2800.7516	ΔΔG	=	181.9852
F	=	348.1440	ΔF	=	0.1005
U	=	359.1960	ΔU	=	0.3624
Fs	=	-3.8525e-4	ΔFs	=	5.4914e-4
Us	=	3.8926e-4	ΔUs	=	1.3997e-4

Figure 3. Comparison between constant and sloped baseline fits. The fit is improved when sloped baselines are included in the fitting model (B). However, limited information in the pre- and post-transition regions results in high uncertainty in the slope of the baselines. Still, the resulting ΔV value is similar to that of a fitting model with constant baseline values (A).

The resulting fits represent the data better than a simple linear 2-state model. It is, however, impossible to distinguish between the two models. For the compressibility model, the error on the compressibility value is as large as the value itself, and the error on ΔV is much greater. For the 3-state model, the values are meaningless in view of the magnitude of the errors. For these higher order models, small experimental error in the dataset will result in large changes in fitted values, which makes it nearly impossible to attain reproducibility. In contrast, because the deviation is reproducible, the parameter values obtained from a linear 2-state fit will also be reproducible (Fig. 5). Often, this required averaging of multiple wavelength scans due to noise in the unfolded baseline.

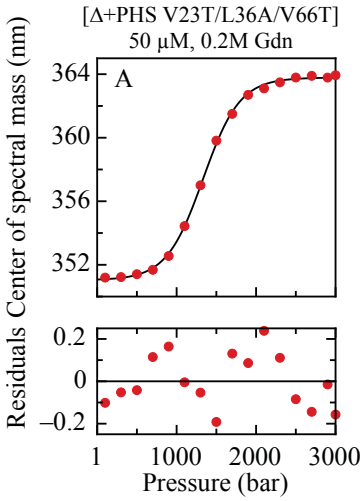


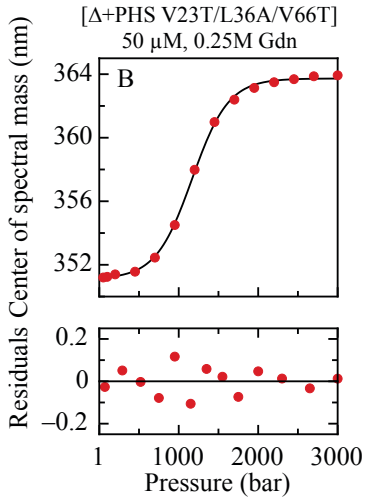


Fitted parameters:			Standard deviations:		
ΔV	=	160.8184	ΔΔV	=	23.8519
ΔG	=	-3557.0657	ΔΔG	=	326.2424
F	=	348.2051	ΔF	=	8.2705e-2
U	=	360.4144	ΔU	=	0.2616
ΔB	=	-5.0052e-2	ΔΔB	=	2.2558e-2

Fitted parameters:			Standard deviations:		
ΔViu	=	101.4704	ΔΔViu	=	94.8053
ΔGiu	=	-2687.4925	ΔΔGiu	=	1034.3358
F	=	348.1111	ΔF	=	8.9497e-2
ΔVni	=	-64.4671	ΔΔVni	=	60.1408
ΔGni	=	-1892.1637	ΔΔGni	=	2920.6564
I	=	355.6368	ΔI	=	12.0448

Figure 4. Comparison between quadratic and 3-state models. The fit is improved when more complex models are used. However, the fitting error becomes too large to extract meaningful values.





Fitted parameters:			Standard deviations:		
ΔV	=	105.7924	ΔΔV	=	3.2685
ΔG	=	-3370.5320	ΔΔG	=	108.9467
F	=	351.0328	ΔF	=	8.9965e-2
U	=	363.7854	ΔU	=	6.3836e-2

Fitted parameters:			Standard deviations:		
ΔV	=	106.7816	ΔΔV	=	3.8068
ΔG	=	-2982.8767	ΔΔG	=	111.1359
F	=	351.1041	ΔF	=	9.0181e-2
U	=	363.7286	ΔU	=	7.2853e-2

Figure 5. Reproducibility of the non-2-state behavior of pressure unfolding profile. The deviation from a simple 2-state model appears to be systematic and an intrinsic characteristic of the pressure unfolding data.

Pressure unfolding monitored by NMR spectroscopy. Uniformly ^{15}N-labeled protein was produced in N-5052 auto-inductive medium (37). Samples were dissolved at 1 mM concentration in 10 mM Tris buffer at pH 7. 10 % of D_2O was added for the lock procedure. Heteronuclear 2D ^{15}N-^{1}H HSQC (38) spectra were recorded on a 600 MHz Bruker Avance^{TM} III instrument spectrometer equipped with a 5 mm Z-gradient ^{1}H-X double-resonance broadband inverse (BBI) probe in a zirconia tube connected to an Xtreme 60 syringe pump (Daedalus Innovations LLC, Philadelphia, PA). In all experiments, the ^{1}H carrier was positioned on the water resonance, and a WATERGATE sequence (39, 40) was incorporated to suppress the solvent signal. The intensities of the backbone amide cross peaks were determined by finding the maximum of a contour spanning the largest value of the peak in a integration box size of 0.3 ppm by 0.03 ppm (^{15}N - ^{1}H) for each peak using the PARIS algorithm (41) included in the GIFA software (42) (Fig. 6).

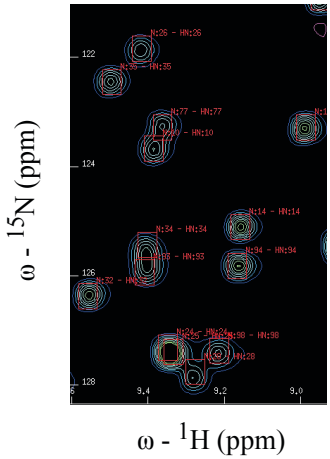


Figure 4. Excerpt of the HSQC spectrum of SNase. The maximum peak intensity for each assigned resonance is found within the marked box.

Amide resonances were assigned at atmospheric pressure from 3D [^{1}H,^{15}N] NOESY-HSQC (mixing time 100 ms) and 3D [^{1}H,^{15}N] TOCSY-HSQC (isotopic mixing 60 ms) double-resonance experiments (43, 44). Equilibrium conditions were evaluated based on the fluorescence pressure jump data and using SoFast HMQC spectra to monitor relaxation (45–48). Native cross-peak intensities were integrated from the corresponding HSQC spectrum and the resulting data points were individually fitted for each resonance, assuming a two-state transition and a linear change of the folding free-energy with pressure (49) (Fig. 7). The peaks were followed until broadened beyond detection, at which point the integrated intensity was equivalent to the background noise of the spectrum. After reaching the maximum pressure, three 1D or 2D spectra were recorded returning to atmospheric pressure to ensure reversibility.

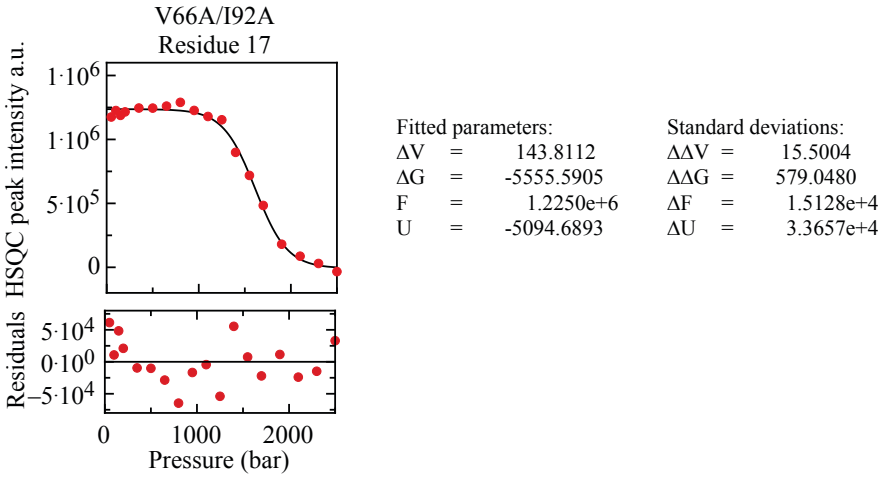


Figure 7. Fitting of a peak intensity profile. Example of a resonance intensity as a function of pressure and a 2-state fitting model.

Almost all peaks in the spectra of V66A/I92A and L25A/V66A/I92A displayed an unexpected increase in intensity before the main transition (Fig. 8). This behavior had not been observed in the parent protein Δ+PHS nor in the single variants, except for a select few amides in I92A. The spectra were checked for peak overlap that might result in this behavior, but only 6 or so peak pairs showed any significant overlap, ruling our the possibility of a spectrum artifact. The effect was still present when tracking the peak volume instead of the maximum peak intensity, suggesting that it is not due to a change in tumbling rates in the system (Fig. 9). This is also supported by the observation that many SNase variants with substitutions far from the protein surface did not show any indication of this phenomenon. Moreover, a similar effect was previously reported in the non-globular Outer Surface protein A (OspA) and using a quartz cell instead of a ceramic one, indicating that it is not a spectroscopic artifact of SNase proteins nor the experimental setup (50). Lastly, the effect was observed in the methyl region of the proton-amide HSQC spectra of SNase variants, indicating that the internal side chains are also reporting on this phenomenon (Fig. 10).

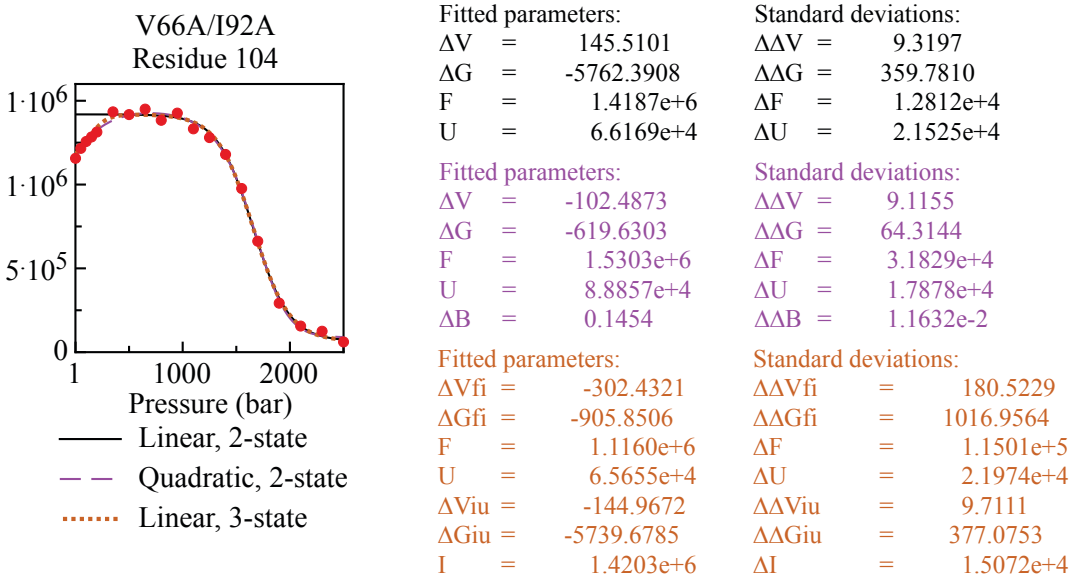


Figure 5. Increase in peak intensity observed before the main transition. Example of a resonance that increases in intensity before the protein unfolds. The intensity reaches a plateau between 500 and 1000 bar that is assumed to be the native baseline for a 2-state model (black) and the points before 500 bar are ignored. The data can be fit well by both a quadratic model that includes a compressibility term (purple) and a 3-state model (orange).

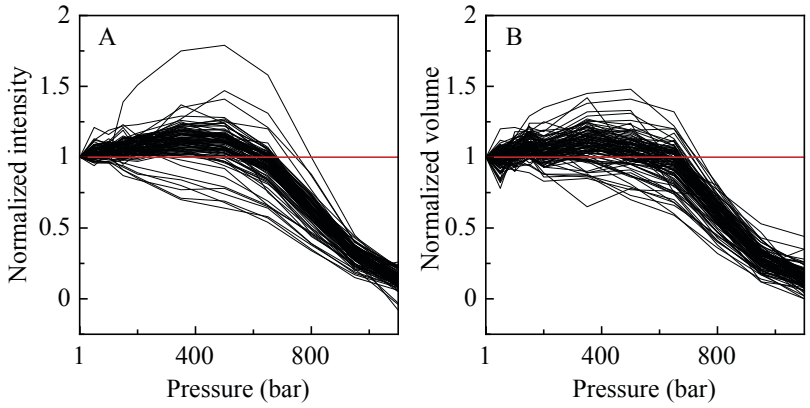


Figure 9. Evidence for a pre-denaturation increase in peak intensity and peak volume. Zoomed in views of the low pressure region of the intensity profile (A) and of the volume profile (B) of amides in L25A/V66A/I92A. The red line denoted the unity line in the normalized profiles.

By 800 bar, the intensity profiles of V66A/I92A reached a clear plateau. A linear, 2-state model was only possible if the plateau value was assumed to represent the folded baseline, and any points before the plateau were ignored (Fig. 8). A quadratic, 2-state model was able to reproduce the data, but the fitted values did not correspond to the expected values, for example, the stabilities obtained independently with Gdn titrations, nor did they have the expected sign. A linear, 3-state model also succeeded in reproducing the data, and in fact returned the same ΔV value as the linear, 2-state fit of the truncated data. However, the error values from the fit of two transitions precluded any insight into the nature of the secondary transition.

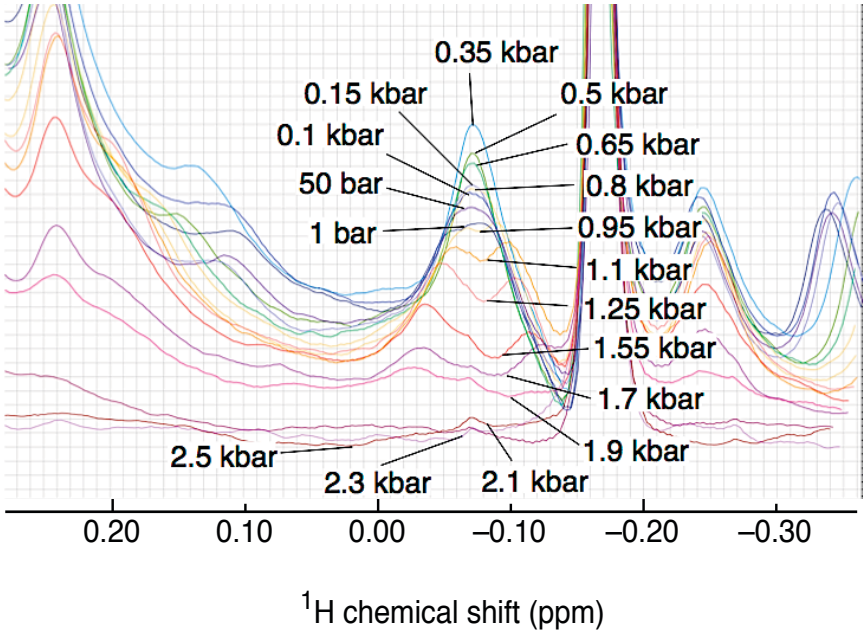


Figure 10. Pre-denaturation increase in peak intensity in the methyls region of the spectrum. 1D proton spectrum near -0.07 ppm. Each line represents a measurement at the specified pressure, with a clear maximum at 350 bar.

When revisiting the previously collected data for I92A (49), the same behavior was found for 21 residues (10, 24, 25, 33, 37, 40, 62, 66, 73, 75, 77, 90, 93, 99, 100, 101, 104, 105, 107, 108 and 129). The spectra were checked for overlapping peaks, which left only 15 residues (37, 62, 66, 73, 75, 77, 90, 93, 99, 100, 101, 104, 105, 107 and 108) showing a pre-transitional increase in intensity. These positions are mostly adjacent but not restricted to the artificial cavity and appear to loosely coincide with sites reporting the largest increases in ΔV as compared to the parent protein (Fig. 11).

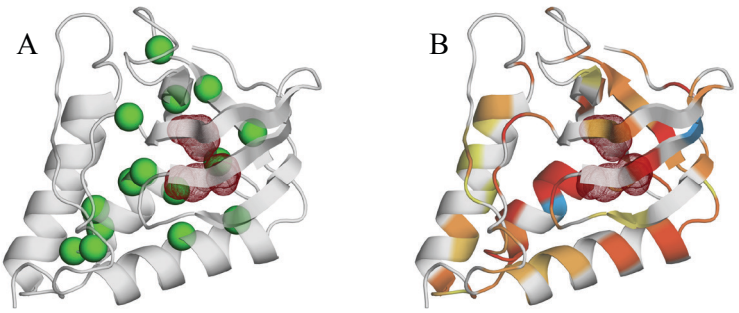


Figure 11. Sites with an initial increase in peak intensity loosely coincide with sites with large changes in ΔV. (A) Mapping of sites in I92A that show a pre-denaturation increase in peak intensity (green spheres). (B) Change in ΔV of I92A with respect to the parent protein Δ+PHS. Increases in ΔV are shown in a scale from yellow to red, with red being the largest increases. Blue sites denote decreases in ΔV. The artificial cavity is shown as red mesh.

The intensity profiles for I92A were markedly different than those for V66A/I92A and the other variants. The intensities appeared to start on a clear plateau and increase only after a certain pressure was reached (about 500 bar) (Fig. 12). They would then reach a maximum at around 800 bar and undergo the main transition. Once again, a linear

3-state model succeeded in reproducing the fitted values of a linear 2-state model of the truncated data, and the initial plateau was assumed to correspond to the native baseline.

Substitution L125A has been observed to increase the cooperative nature of SNase under pressure (51), and has been suggested to populate an intermediate state as evidenced by data-constrained MD simulations (49). It is interesting to note that the increase in peak intensity observed in the variants I92A, V66A/I92A and L25A/V66A/I92A becomes less pronounced in variant I92A/L125A. This is consistent with a highly local destabilization that results from artificial cavities that biases the native state ensemble and its unfolding routes as pressure increases (48).

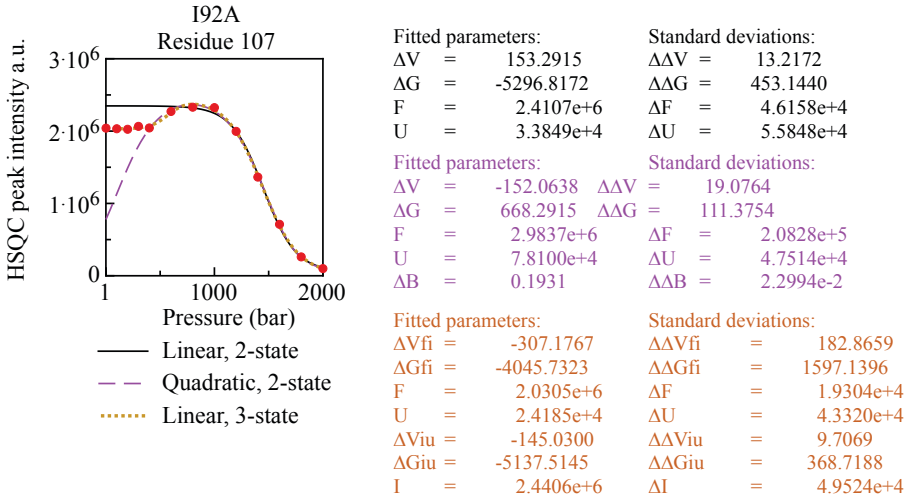


Figure 12. Increase in intensity occurs only after 500 bar. Unlike the curves observed for V66A/I92A, the 15 cross peaks of I92A showed a delayed increase in intensity starting only after about 500 bar. A linear 2-state model cannot reproduce the curve (black). Instead, the fit ignored the initial data points until a maximum in intensity is reached. A quadratic model cannot reproduce the curve either (purple). Instead, the fit ignored the initial data points before the increase in intensity (about 500 bar). A 3-state model successfully reproduces the fit, and yields values with low error for all three baselines (orange).

It is unlikely that this phenomenon reports on a discrete intermediate state that exists in slow exchange with the native state of SNase. No evidence for a well-defined intermediate state can be observed in the high-pressure fluorescence (using the type of analysis in Fig 4B), or in the HSQC spectra. Instead, a possible interpretation of such a phenomenon is the increase in population as a function of pressure of the species responsible for the resonance, and a decrease in the heterogeneity of the population ensemble. This implies that the native state ensemble at 1 bar involves species of various molar volumes in fast exchange with each other. Those of higher volume become depleted as pressure increases, and the specie(s) of lowest volume become(s) dominant. This is consistent with direct observation of the compression of the folded structure of proteins (52, 53, 35).

Bibliography

1. Dwyer JJ, Gittis AG, Karp DA, Lattman EE, Spencer DS, Stites WE, García-Moreno E. B (2000) High apparent dielectric constants in the interior of a protein reflect water penetration. *Biophys. J.* 79(3):1610–1620.

2. Karp DA, Gittis AG, Stahley MR, Fitch CA, Stites WE, García-Moreno E. B (2007) High apparent dielectric constant inside a protein reflects structural reorganization coupled to the ionization of an internal Asp. *Biophys. J.* 92(6):2041–2053.

3. Shortle D, Meeker AK (1989) Residual structure in large fragments of staphylococcal nuclease: effects of amino acid substitutions. *Biochemistry* 28(3):936–944.

4. Bertani G (1951) Studies on lysogenesis I. The mode of phage liberation by lysogenic escherichia coli. *J. Bacteriol.* 62(3):293–300.

5. Rosenberg AH, Lade BN, Dao-shan C, Lin SW, Dunn JJ, Studier FW (1987) Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene* 56(1):125–135.

6. Studier FW, Moffattf BA (1986) Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-level Expression of Cloned Genes. *J. Mol. Biol.* 189(1):113–130.

7. Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW (1990) Use of T7 RNA Polymerase to Direct Expression of Cloned Genes. *Methods Enzymol.* 185:60–89.

8. Fuchs S, Cuatrecasas P, Anfinsen B (1967) An Improved Method for the Purification of Staphylococcal Nuclease. *J. Biol. Chem.* 242(20):4768–4770.

9. Sambrook J, Russell DW (2001) *Molecular cloning* (Cold Spring Harbor Laboratory Press, New York). 3rd. Ed.

10. Rhodes G (2006) *Crystallography made crystal clear* (Academic Press, Burlington). 3rd. Ed.

11. Chimenti MS, Khangulov VS, Robinson AC, Heroux A, Majumdar A, Schlessman JL, García-Moreno E. B (2012) Structural reorganization triggered by charging of Lys residues in the hydrophobic interior of a protein. *Structure* 20(6):1071–85.

12. Truckses DM, Somoza JR, Prehoda KE, Miller SC, Markley JL (1996) Coupling between trans/cis proline isomerization and protein stability in staphylococcal nuclease. *Protein Sci.* 5:1907–1916.

13. Otwinowski Z, Minor W (1997) in *Methods in Enzymology, Vol 276: Macromolecular Crystallography, part A*, eds Carter Jr. CW, Sweet RM (Academic Press (New York)), p 307-326.

14. McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr. Sect. D - Biol. Crystallogr.* 61(4):458–464.

15. Collaborative Computational Project N 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. Sect. D - Biol. Crystallogr.* 50(5):760–763.

16. Krissinel EB, Winn MD, Ballard CC, Ashton AW, Patel P, Potterton E a, McNicholas SJ, Cowtan KD, Emsley P (2004) The new CCP4 Coordinate Library as a toolkit for the design of coordinate-related applications in protein crystallography. *Acta Crystallogr. Sect. D - Biol. Crystallogr.* 60(12):2250–2255.

17. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D - Biol. Crystallogr.* 53(3):240–255.

18. Vaguine AA, Richelle J, Wodak SJ (1999) SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr. Sect. D - Biol. Crystallogr.* 55(1):191–205.

19. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D - Biol. Crystallogr.* 66(1):12–21.

20. Painter J, Merritt EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. Sect. D - Biol. Crystallogr.* 62(4):439–450.

21. Kabsch W (1976) A solution for best rotation to relate 2 sets of vectors. *Acta Crystallogr. Sect. A* 32(5):922–923.

22. Hess B, Van Der Spoel D, Lindahl E (2008) GROMACS 4 : Algorithms for Highly Efficient , Load-Balanced , and Scalable Molecular Simulation. *J. Chem. Theory Comput.* 4(3):435–447.

23. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. 725(May):712–725.

24. Till MS, Ullmann GM (2010) McVol - a program for calculating protein volumes and identifying cavities by a Monte Carlo algorithm. *J. Mol. Model.* 16(3):419–429.

25. Santoro MM, Bolen DW (1988) Unfolding Free Energy Changes Determined by the Linear Extrapolation Method . 1. Unfolding of Phenylmethanesulfonyl a-Chymotrypsin Using Different Denaturants. *Biochemistry* 27:8063–8068.

26. Shortle D, Stites WE, Meeker AK (1990) Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease. *Biochemistry* 29(35):8033–41.

27. Whitten ST, García-Moreno E. B (2000) pH Dependence of Stability of Staphylococcal Nuclease: Evidence of Substantial Electrostatic Interactions in the Denatured State. *Biochemistry* 39(46):14292–14304.

28. Pace CN (1986) Determination and Analysis of Urea and Guanidine Hydrochloride Denaturation Curves. *Methods Enzymol.* 131:266–280.

29. Bouvignies G, Vallurupalli P, Hansen DF, Correia BE, Lange O, Bah A, Vernon RM, Dahlquist FW, Baker D, Kay LE (2011) Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* 477(7362):111–114.

30. Isom DG, Castañeda CA, Cannon BR, García-Moreno E. B (2011) Large shifts in pK a values of lysine residues buried inside a protein. *Proc. Natl. Acad. Sci. U.S.A.* 108(13):5260–5265.

31. Isom DG, Castañeda CA, Cannon BR, Velu P, García-Moreno E. B (2010) Charges in the hydrophobic interior of proteins. *Proc. Natl. Acad. Sci. U.S.A.* 107(37):16096–16100.

32. Isom DG, Cannon BR, Castañeda CA, Robinson A, García-Moreno E. B (2008) High tolerance for ionizable residues in the hydrophobic interior of proteins. *Proc. Natl. Acad. Sci. U.S.A.* 105(46):17784–8.

33. Shortle D, Meeker AK (1986) Mutant forms of staphylococcal nuclease with altered patterns of guanidine hydrochloride and urea denaturation. *Proteins Struct. Funct. Bioinforma.* 1(1):81–9.

34. Maeno A, Sindhikara D, Hirata F, Otten R, Dahlquist FW, Yokoyama S, Akasaka K, Mulder FAA, Kitahara R (2015) Cavity as a Source of Conformational Fluctuation and High-Energy State : High-Pressure NMR Study of a Cavity-Enlarged Mutant of T4Lysozyme. *Biophysj* 108:133–145.

35. Nucci NV, Fuglestad B, Athanasoula EA, Wand AJ (2014) Role of cavities and hydration in the pressure unfolding of T4 lysozyme. *Proc. Natl. Acad. Sci. U.S.A.* 111(38):13846–13851.

36. Dellarole M, Royer CA (2014) in *Fluorescence spectroscopy and microscopy*, Methods in Molecular Biology. (Humana Press), pp 53–74.

37. Studier FW (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* 41(1):207–234.

38. Bodenhausen G, Ruben DJ (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* 69(1):185–189.

39. Piotto M, Saudek V, Sklenár V (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* 2(6):661–665.

40. Sklenar V (1991) in *NMR Applications in Biopolymers* (Springer), pp 63–84.

41. Stoven V, Mikou A, Lallemand J-Y (1989) PARIS, a program for automatic recognition and integration of 2D NMR signals. *J. Magn. Reson.* 82(1):163–168.

42. Pons JL, Malliavin TE, Delsuc M a (1996) Gifa V. 4: A complete package for NMR data set processing. *J. Biomol. NMR* 8(4):445–452.

43. Bax A, Pochapsky SS (1992) Optimized Recording of Heteronuclear Multidimensional NMR Spectra Using Pulsed Field Gradients. *J. Magn. Reson.* 99(3):638–643.

44. Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn AM, Clore GM (1989) Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to . *Biochemistry* 28(15):6150–6.

45. Schanda P, Kupce E, Brutscher B (2005) SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J. Biomol. NMR* 33(4):199–211.

46. Gal M, Schanda P, Brutscher B, Frydman L (2007) UltraSOFAST HMQC NMR and the repetitive acquisition of 2D protein spectra at Hz rates. *J. Am. Chem. Soc.* 129(5):1372–7.

47. Gal M, Kern T, Schanda P, Frydman L, Brutscher B (2009) An improved ultrafast 2D NMR experiment: towards atom-resolved real-time studies of protein kinetics at multi-Hz rates. *J. Biomol. NMR* 43(1):1–10.

48. Roche J, Dellarole M, Caro JA, Norberto DR, Garcia AE, García-Moreno E. B, Roumestand C, Royer CA (2013) Effect of Internal Cavities on Folding Rates and Routes Revealed by Real-time Pressure-Jump NMR Spectroscopy. *J. Am. Chem. Soc.* 82(6):1069–1080.

49. Roche J, Caro JA, Norberto DR, Barthe P, Roumestand C, Schlessman JL, Garcia AE, García-Moreno E. B, Royer CA (2012) Cavities determine the pressure unfolding of proteins. *Proc. Natl. Acad. Sci. U.S.A.* 109(18):6945–6950.

50. Kitahara R, Simorellis AK, Hata K, Maeno A, Yokoyama S, Koide S, Akasaka K (2012) A delicate interplay of structure, dynamics, and thermodynamics for function: a high pressure NMR study of outer surface protein A. *Biophys. J.* 102(4):916–926.

51. Roche J, Dellarole M, Caro JA, Guca E, Norberto DR, Yang Y, Garcia AE, Roumestand C, García-Moreno E. B, Royer CA (2012) Remodeling of the Folding Free Energy Landscape of Staphylococcal Nuclease by Cavity-Creating Mutations. *Biochemistry* 51(47):9535–9546.

52. Nagae T, Kawamura T, Chavas LMG, Niwa K, Hasegawa M, Kato C, Watanabe N (2012) High-pressure-induced water penetration into 3-isopropylmalate dehydrogenase. *Acta Crystallogr. Sect. D - Biol. Crystallogr.* 68(3):300–309.

53. Day R, Garcia AE (2008) Water penetration in the low and high pressure native states of ubiquitin. *Proteins Struct. Funct. Bioinforma.* 70(4):1175–1184.

APPENDIX:

STRUCTURAL DETERMINANTS OF THE CONFORMATION OF INTERNAL IONIZABLE RESIDUES IN PROTEINS

Introduction

Charges are incompatible with hydrophobic environments. Proteins generally display ionizable residues on the surface, where they can interact with bulk water (1). However, many of the biological processes that are essential for life such as energy transduction involve proteins that display ionizable residues buried in the hydrophobic core that are indispensable for function (2–6). The structural and energetic consequences of burying a charge were explored systematically in a survey of over 100 variants of SNase with an internal ionizable residue (7–10). The pK_a value of the substituted ionizable residue experienced a shift from the value for the model compound amino acid that ranged from no detectable shift to a shift greater than 5 pH units, equivalent to an energetic penalty of about 7 kcal/mol. The magnitude of the shift, and of the destabilization suffered by the protein upon burial of an ionizable residue, depended on where within the protein interior the substitution was made.

Standard, structure-based computational methods fail to accurately predict the experimentally measured pK_a values of ionizable residues in the protein interior because they do not correctly account for the factors that determine the dielectric relaxation of the protein (11, 12). Hydration and depth of burial, polarity and polarizability, and conformational entropy and local structural fluctuations are all factors that can contribute to the heterogeneous dielectric environment experienced by the ionizable residue buried within the interior of a protein (11). Crystal structures of over 30 variants with internal ionizable residues substituted at various internal positions showed that the ionizable moiety was indeed buried in the protein core (11, 13–19). The structures showed how some the observed conformation of the ionizable side chain could be explained by its

proximity to polar contacts with backbone and other side chains. Internal cavities also appeared to influence the side chain conformation. When within reach, the side chain adopted a conformation that allowed it to occupy the natural cavity observed in the parent protein. Alternatively, internal waters were sometimes observed to fill the cavity and hydrogen bond to the side chain (13). In some cases, however, the side chain adopted a conformation with no polar contacts within 5 Å of the ionizable moiety (13, 17). These cases corresponded to the largest measured shifts in pK_a of the ionizable residue and the highest energetic penalties to the protein (7, 8). Still, the side chain often adopted a discrete and well-defined rotameric state. Here, the determinants of the conformation of internal ionizable residues were examined systematically and with structural detail by engineering polar substitutions and cavities in the vicinity of internal Lys residues.

In variants with an internal Lys, a substitution was engineered in an attempt to alter its conformation and bias the side chain toward a new microenvironment. The substitutions introduced a nearby polar residue, introduced a nearby artificial cavity, or eliminated a nearby cavity. Crystal structures were solved to inspect the changes in the conformation and microenvironment of the buried Lys.

This work shows how modulating the polarity and cavity volume around a buried ionizable residue directly determines the conformation of the side chain. Eventually, thermodynamic stability measurements as a function of pH will report on the energetic consequences of altering the conformational state of the buried Lys and allow us to relate the conformational changes observed in crystal structures with the energetic consequences and the pK_a values of internal ionizable residues. This work will provide the experimental data that will help develop accurate computational methods for

prediction of pK_a values and electrostatic forces. It also provides a rationale for the modulation of the microenvironment of functional residues in efforts to engineer novel or altered enzymatic active sites in proteins.

Results

Crystal structures. The structure of I92K was solved to 1.85 Å resolution in 2004 (17) but the structure factors were not deposited (PDB ID 1TT2). In this structure, the side chain of Lys-92 was observed to adopt two conformers with equal occupancy. Neither conformer reached the natural cavity nor made any polar contacts. The variant was re-crystallized for this study, yielding a structure of 1.75 Å (Fig 1A). In this structure, the side adopted a single conformer that matched up to the Cε the main conformer of structure 1TT2. The Cε to Nζ bond, however, pointed in the opposite direction. Some evidence of a second conformer could be observed in the electron density, but when modeled, was not corroborated by the electron density. Faint density was observed in both 2Fo-Fc and Fo-Fc maps disconnected from the density at 92 and in a position where a buried water molecule could be observed in other structures of SNase with a buried polar or ionizable residue. A 0.5 occupancy water molecule yielded good agreement with the density maps and a B-factor comparable to those of Lys-92. The structure of V23K was solved previously (RCSB ID: 3QOJ) (Fig 1D). The buried Lys residue was observed to adopt a single conformation in both structures. Lys-23 was positioned in the natural cavity observed in the parent structure. There were no polar groups interacting directly with the Nζ, although Thr-62, an interfacial water molecule and the carbonyl of Gly-20 were within 5 Å of it.

An artificial cavity was engineered next to the buried Lys by substituting a large hydrophobic residue to Ala. In the case of I92K, Leu-25 was truncated to Ala, yielding variant L25A/I92K. Position 25 is directly across the β-barrel of position 92 and the side chains project toward each other. The structure of the variant L25A/I92K showed that the conformation of Lys-92 changed, with a rotation about the Cβ to point directly at Ala-25 (Fig. 1C). There were no polar groups within 5 Å of the Nζ.

In contrast, when Ala-92 is introduced to variant V23K, the structure of V23K/I92A showed effectively no change in the conformation of Lys-23. Despite the large artificial cavity generated at position 23, the Nζ of Lys-23 continued to occupy the natural cavity of SNase and to be in the vicinity (<5 Å) of 3 polar groups.

The previously published structure of L25K showed Lys-25 completely surrounded by hydrophobic side chains, too far from the natural cavity to occupy it (Fig. 1G). Variant L25K/I92F was engineered to sterically hinder the conformation of Lys-25. The structure indeed showed a different conformer for Lys-25, in which the side chain is exposed to the surface of the protein. In order to achieve this, the protein underwent a structural rearrangement that shifted β-strands 1, 2 and 3 away from their positions in the parent protein by more than 2 Å (Fig. 1H). The new microenvironment of the Lys side chain consisted of three polar contacts within 3.5 Å to the side chain and carbonyl of Thr-13 and the carbonyl of Lys-26.

Variant L25K/I92T was engineered, with the polar side chain Thr at position 92. The structure of this double variant shows the Lys side chain in an identical conformation to the one observed for variant L25K, except it is making a hydrogen bond with Thr-92, which in turn is hydrogen bonded to a single, full occupancy water molecule that also

contacts the carbonyl of Ala-90 (Fig. 2B and C). The site where the water molecule was found was previously occluded by the side chain of Ile-92 and is equivalent to waters found at site 6 in the previously used nomenclature of Schlessman et al.(13).

Bibliography

1. Perutz MF, Kendrew JC, Watson HC (1965) Structure and function of haemoglobin 2. Some relations between polypeptide chain configuration and amino acid sequence. *J. Mol. Biol.* 13(3):669–678.

2. Warshel A (1978) Energetics of enzyme catalysis. *Proc. Natl. Acad. Sci. U.S.A.* 75(11):5250–5254.

3. Cannon WR, Benkovic SJ (1998) Solvation, Reorganization, Energy , and Biological Catalysis *. *J. Biol. Chem.* 273(41):26257–26260.

4. Von Ballmoos C, Wiedenmann A, Dimroth P (2009) Essentials for ATP synthesis by F1F0 ATP synthases. *Annu. Rev. Biochem.* 78:649–72.

5. Harris TK, Turner GJ (2002) Structural Basis of Perturbed pK a Values of Catalytic Groups in Enzyme Active Sites. *IUBMB Life* 53:85–98.

6. Iwata S, Lee JW, Okada K, Lee JK, Iwata M, Rasmussen B, Link TA, Ramaswamy S, Jap BK (1998) Complete Structure of the 11-Subunit Bovine Mitochondrial Cytochrome bc 1 Complex. 281(July):64–71.

7. Isom DG, Castañeda CA, Cannon BR, García-Moreno E. B (2011) Large shifts in pK a values of lysine residues buried inside a protein. *Proc. Natl. Acad. Sci. U.S.A.* 108(13):5260–5265.

8. Isom DG, Castañeda CA, Cannon BR, Velu P, García-Moreno E. B (2010) Charges in the hydrophobic interior of proteins. *Proc. Natl. Acad. Sci. U.S.A.* 107(37):16096–16100.

9. Isom DG, Cannon BR, Castañeda CA, Robinson A, García-Moreno E. B (2008) High tolerance for ionizable residues in the hydrophobic interior of proteins. *Proc. Natl. Acad. Sci. U.S.A.* 105(46):17784–8.

10. Harms MJ, Schlessman JL, Sue GR, García-Moreno E. B (2011) Arginine residues at internal positions in a protein are always charged. *Proc. Natl. Acad. Sci. U.S.A.* 108(47):18954–9.

11. García-Moreno E. B, Dwyer JJ, Gittis AG, Lattman EE, Spencer DS, Stites WE (1997) Experimental measurement of the effective dielectric in the hydrophobic core of a protein. *Biophys. Chem.* 64(1-3):211–224.

12. Fitch CA, Karp DA, Lee KK, Stites WE, Lattman EE, García-Moreno E. B (2002) Experimental pKa Values of Buried Residues: Analysis with Continuum Methods and Role of Water Penetration. *Biophys. J.* 82(6):3289–3304.

13. Schlessman JL, Abe C, Gittis A, Karp DA, Dolan MA, García-Moreno E. B (2008) Crystallographic study of hydration of an internal cavity in engineered proteins with buried polar or ionizable groups. *Biophys. J.* 94(8):3208–16.

14. Dwyer JJ, Gittis AG, Karp DA, Lattman EE, Spencer DS, Stites WE, García-Moreno E. B (2000) High apparent dielectric constants in the interior of a protein reflect water penetration. *Biophys. J.* 79(3):1610–1620.

15. Karp DA, Gittis AG, Stahley MR, Fitch CA, Stites WE, García-Moreno E. B (2007) High apparent dielectric constant inside a protein reflects structural reorganization coupled to the ionization of an internal Asp. *Biophys. J.* 92(6):2041–2053.

16. Stites WE, Gittis a G, Lattman EE, Shortle D (1991) In a staphylococcal nuclease mutant the side-chain of a lysine replacing valine 66 is fully buried in the hydrophobic core. *J. Mol. Biol.* 221(1):7–14.

17. Nguyen DM, Leila Reynald R, Gittis AG, Lattman EE (2004) X-ray and thermodynamic studies of staphylococcal nuclease variants I92E and I92K: insights into polarity of the protein interior. *J. Mol. Biol.* 341(2):565–74.

18. Harms MJ, Castañeda CA, Schlessman JL, Sue GR, Isom DG, Cannon BR, García-Moreno E. B (2009) The pK(a) values of acidic and basic residues buried at the same internal location in a protein are governed by different factors. *J. Mol. Biol.* 389(1):34–47.

19. Harms MJ, Schlessman JL, Chimenti MS, García-Moreno E. B, Sue GR, Damjanovic A (2008) A buried lysine that titrates with a normal pKa : Role of conformational flexibility at the protein – water interface as a determinant of pKa values. *Protein Sci.* 17:833–845.

Figure 1. (A) Structure of I92K. Side chain of Lys-92 is shown as sticks and cavities are shown as red mesh. (B) Structure of L25A/I92K. Side chains of Ala-25 and Lys-92 are shown as sticks. (C) Microenvironment of Lys-92 in structure L25A/I92K, highlighting all positions within 5 Å of the ionizable moiety Nζ of Lys-92. (D) Structure of V23K. (E) Structure of V23K/I92A. (F) Microenvironment of Lys-23 in structure V23K/I92A, highlighting all positions within 5 Å of the ionizable moiety Nζ of Lys-23. Interfacial water molecule 155 is at 4.7 Å and highlighted as a blue sphere. (G) Structure of L25K. (H) Structure of L25K/I92F. The backbone is traced in blue wire and compared to the trace of L25K in red wire. (I) Microenvironment of Lys-25 in structure L25K/I92F, highlighting all positions within 5 Å of the ionizable moiety Nζ of the partially exposed Lys-25. Two water molecules of the hydration layer are shown.

Figure 2. (A) Structure of L25K. (B) Structure of L25K/I92T, with an internal water molecule shown as a blue sphere. (C) Microenvironment of Lys-25 in structure V23K/I92A, highlighting all positions within 5 Å of the ionizable moiety Nζ of Lys-25. The internal water molecule contacts the carbonyl of Ala-90 and is 2.7 Å from Thr-92 and 4 Å from Lys-25. Thr-92 hydrogen bonds to Lys-25 at 2.9 Å.

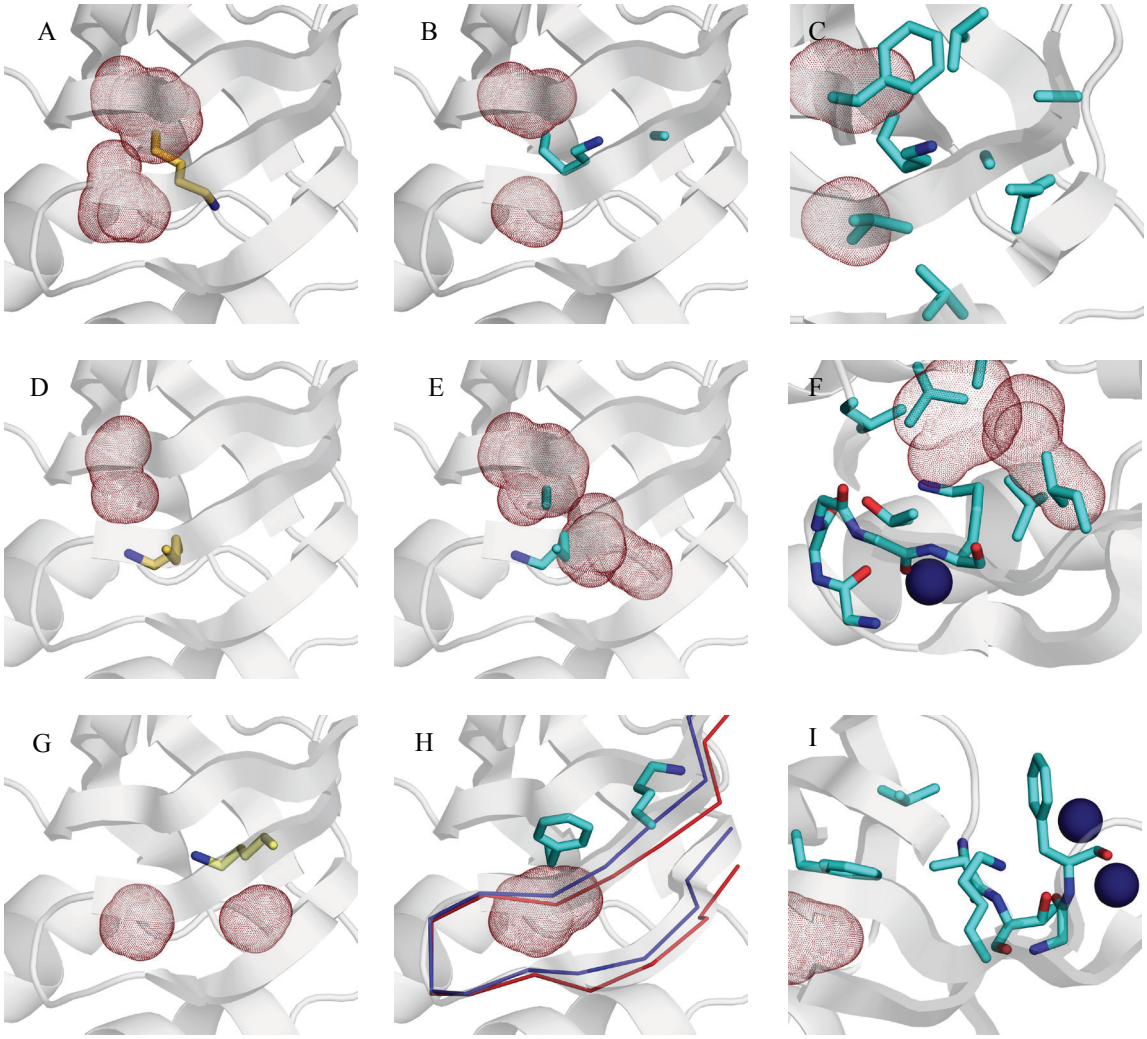


Figure 1.

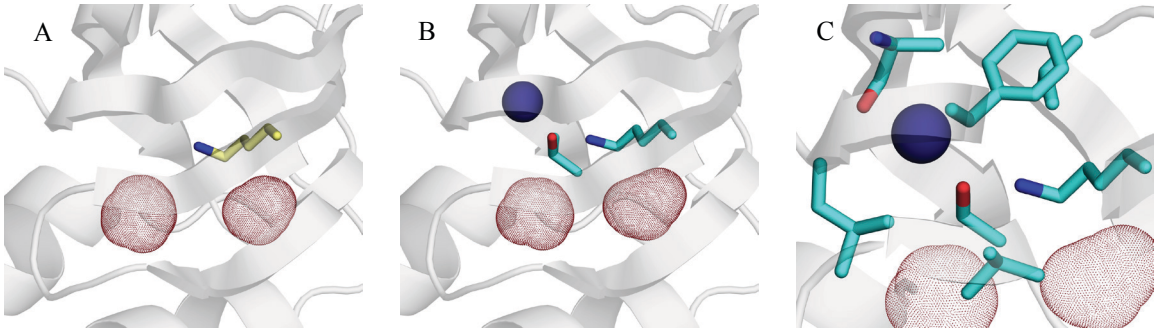


Figure 2.

Vita

José Alfredo Caro was born in San Carlos de Bariloche, Argentina in 1985 to José Alfredo Caro and Magdalena Serrano de Caro. He is the third generation to carry this name, which honors his grandfather – the Argentinean botanist. Alfredo lived but a few years in the foothills of the Andes before his parents brought him and his older sister Florencia to Switzerland in light of Argentina's economic turmoil. There, he became fluent in Swiss German. His family returned to Bariloche in 1993, just in time for his 9th birthday. Alfredo attended a bilingual school that would help the transition to an essentially new language to him, Spanish. He played catch-up to understand the culture where he was born. Aware of alternatives, and particularly at odds with the inflexible educational system in Argentina, Alfredo sought out higher education in the United States, and was accepted, through a late appeal, to the University of California, Santa Barbara's Department of Physics in 2003. His decision to pursue physics was influenced by his parents, both Doctors in Condensed Matter Physics. The escapism this métier provided his parents, he decided, would allow him to explore outside the bounds of country, culture and language.

With no background in Biology but a sense of heritage, Alfredo began work on a Ph.D. in Biophysics at Johns Hopkins University. His work in the laboratory of Dr. Bertrand García-Moreno Esteva concluded in February of 2015. In that time, Alfredo played semi-professional soccer in the Maryland Majors top division, became a U.S. citizen, and survived forty days and forty nights of experiments in southern France. Alfredo presently works in the laboratory of Dr. Joshua Wand at the University of Pennsylvania in Philadelphia.